HD

Structural Studies of Proteins Using Computational Methods

Thesis for the Degree of
Doctor of Philosophy

Sami Raza
Department of Chemistry
University of Leicester

May 1998

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

Protein function is inextricably linked with protein three-dimensional (3D) structure. Therefore, a greater understanding of the 3D structure of proteins leads to a deeper understanding of how and why a particular protein functions in a specific way.

Computer-based methods have been used to study the relationship between protein structure and function. The methods applied to the work presented in this thesis are: structure prediction through comparative modelling, both by homology (*i.e.* sequence similarity) with, and by analogy (*i.e.* threading) to, proteins with known 3D structure; structure generation using distance geometry and simulated annealing calculations on NMR derived data; assessment of generated structures; and molecular interaction of 3D structures, particularly studies on electrostatic surface potential, ligand binding and molecular docking. These modelling methods have been applied as follows:

- Comparative modelling of human NADPH cytochrome P_{450} reductase, including an investigation into the interaction of its component domains. Specific residues were identified as possibly being involved in electron transfer from the FAD to the FMN prosthetic groups within the enzyme.

- Comparative modelling of ubiquitin conjugating enzyme 9, glutathione transferase and neurocalcin delta for the purpose of the 1996 CASP2 modelling assessment. The models compared favourably with models from other contributing groups, and predicted well the errors in the atomic positions.

- The solution structure of the GDP-bound form of the G-protein, Cdc42Hs, was determined using experimentally derived NMR data. Chemical shift changes (obtained by collaborators) between the GTP-bound and GDP-bound forms indicate that conformational change between the two states, facilitating interaction with effector and regulatory proteins, are associated with a contiguous region on the surface of the protein.

- Docking a ligand into the crystal structure of the third PDZ domain of the human homologue of *Drosophila* discs-large tumour suppressor protein; thus providing a structural hypothesis for ligand specificity.

- Modelling the role of the enzyme trimethylamine dehydrogenase in electron transfer, substrate access and substrate specificity. Channels characterised within the enzyme indicate possible additional regions for substrate access to the binding site. A number of residues were identified as having possible roles in specificity for trimethylamine over dimethylamine as substrate.

STATEMENT

The research presented in this thesis was performed by the author in the Department of Chemistry, University of Leicester, UK, between 1994 and 1997. The work has not been submitted for any other degree at this or any other university.

Signed: Sami Raza Date: 29 May 1998

Sami Raza
Department of Chemistry
University of Leicester
University Road
Leicester LE1 7RH
UK

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr Michael J Sutcliffe, for his patience and support throughout the duration of these studies.

Thanks also to collaborators: Prof. Gordon Roberts, Dr. Sandeep Modi, Dr. Geoff Briggs, Prof. Roland Wolf, Dr. Mark Paine, Prof. Robert Oswald, Dr. Joanna Feltham, Prof. Robert Liddington, Dr. Joao Cabral, Dr. Nigel Scrutton and Peter Roberts.

Finally, I wish to thank the University of Leicester for the research scholarship.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Proteins

Proteins are essential to the workings of living organisms and are involved in a diverse range of biological functions. For example, enzymes are proteins that catalyse chemical reactions (*e.g.* digestion); they are also involved in protein regulation, that is, the control and translation of DNA into protein within the cell. Proteins are involved in messenger pathways (*e.g.* acting as receptors in nerve cell junctions). They may function as part of the immune system, as antibodies. Proteins may also perform the functions of transport (*e.g.* transport of oxygen by haemoglobin), storage (*e.g.* iron is stored in the liver as a complex with ferritin) or be involved in structural support (*e.g.* collagen).

Naturally-occurring proteins are made up of sequences composed of 20 different amino acids, which may be categorised according to the chemical nature of their sidechains (*e.g.* acidic, basic, aliphatic, aromatic, *etc*). The positions of the amino acids within the protein determine both the protein's three-dimensional (3D) structural and chemical properties.

Protein structure may be classified into four levels: primary, secondary, tertiary and quaternary. Primary structure refers to the linear sequence of amino acids that comprises a particular protein. Secondary structure refers to the folding of some of the sequence into regular structures (*e.g.* α-helices, β-sheets) held together by hydrogen bonds. Tertiary structure refers to the spatial arrangement of the secondary structural elements relative to each other and of the irregular structural regions between them; that is, the formation of protein domains or folds. Quaternary structure relates to protein molecules which consist of more than one polypeptide chain, and refers to the spatial arrangement of these separate chains with respect to each other.

The function of a protein is inextricably linked with its 3D structure. Knowledge of a protein's 3D structure is, therefore, vitally important in acquiring a more complete understanding of how that protein functions in its biological context.

1.2 Experiment-Based Approaches to Protein 3D Structure Determination

The 3D structures of proteins may be determined experimentally using X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy and/or Electron Microscopy (EM).

In X-ray crystallography (see, for example, Rhodes, 1993), the protein is prepared in crystal form, and bombarded with X-ray radiation which is diffracted by the protein's electrons in the symmetrically packed crystal to produce a scattering pattern. The diffracted waves, defined by their amplitude and phase, can be used to calculate the electron density of the protein, and thus a model of its 3D structure. The amplitudes of diffracted waves can be calculated directly from the intensities of the peaks in the scattering pattern. However, determining the phases, or relative shifts, of the diffracted waves with respect to each other is a non-trivial problem since they cannot be calculated directly from the scattering pattern but must be deduced mathematically.

NMR spectroscopy (see, for example, Slichter, 1990), performed on proteins in solution, is based on the existence of a magnetic moment in certain types of atomic nucleus (*e.g.* ^{1}H, ^{13}C, ^{15}N) and its resonance, or spin alignment, with an externally applied magnetic field. However, the magnetic field experienced by a particular proton, for example, undergoes slight perturbation due to the small magnetic fields generated by the protons of neighbouring atoms. This perturbation, or chemical shift, is characteristic of the environment of that particular proton. Knowledge of the correlation between chemical shifts and atomic environment is thus used for structural identification purposes. For a given protein molecule, a multi-dimensional (*e.g.* 2D) NMR spectrum may be generated, consisting of cross-peaks that indicate interactions between particular nuclei. These interactions can either be through space or through bonds. The intensities of the cross-peaks corresponding to these interactions can be translated into distance restraints (from the through-space interactions) and dihedral angle restraints (from the through-bond interactions). Once these restraints are assigned to particular atoms, an ensemble of 3D models, each of which satisfies the NMR distance and dihedral angle restraints, may be calculated using distance geometry techniques followed by further refinement of the model using simulated annealing.

For both X-ray crystallography and NMR techniques, the structure determination for a protein may not be possible at all because, for example, the protein cannot be prepared in crystal form (X-ray crystallography) or in solution (NMR), or perhaps great difficulty is experienced in solving the phase problem (X-ray crystallography), or the protein may be too large, that is, greater than approximately 250 residues (NMR). The 3D structure determination of membrane proteins and cytoskeletal proteins and viruses is an area of focus for EM, sometimes allowing the difficulties experienced by X-ray/NMR techniques to be overcome.

In EM (see, for example, Haggis, 1966), an electron beam is applied to a thin specimen, generating image contrast (areas of light and dark) which is recorded on photographic film or an electronic detector. A series of projections of the specimen taken at various tilt angles can be combined to produce a 3D image. Developments in EM and image processing allow 3D images of biological samples to approach, but not quite achieve, atomic resolution.

Biological materials are composed of relatively light chemical elements with similar electron scattering amplitudes, and therefore generate little image contrast. Heavy metal stains are employed to increase contrast and to preserve the specimen in dehydrated form. However, drying the sample exposes it to high concentrations of salts and heavy metal stain, and may cause damage to the specimen.

One method used to avoid dehydration and staining problems is cryo-EM (Dubrochet *et al.*, 1988). A thin suspension of the specimen is rapidly cooled, avoiding water crystallisation and the associated specimen damage, the sample becoming embedded in a glass-like solid (vitreous ice). The unstained sample is observed at low temperature, and remains in hydrated form.

1.3 Prediction-Based Approaches to Protein 3D Structure Generation

Currently, the number of proteins with experimentally determined (X-ray crystallography / NMR), or 'known', 3D structure stands at approximately 5000, and this constitutes only a small percentage of the number of proteins with known amino acid sequence, approximately 200 000, highlighting the existence of a significant information gap. Approximately 30% of known protein sequences have at least 25% residue identity with one of the entries in the Brookhaven Protein Data Bank (PDB; Abola *et al.*, 1987), indicating that the scope of application for modelling on the basis of sequence similarity is favourable (Chothia, 1992). That is, the 3D structure could be modelled for approximately 30% of the proteins with known sequence, which is ten times the number of proteins with experimentally determined 3D structures at the present time.

However, the 3D structure of a query protein can be modelled based not only on its homology (comparability based on similarity arising from divergent evolution) to, but also on its analogy (comparability based on similarity outside the scope of divergent evolution) with, proteins that have known 3D structure. Structural homology is exhibited between proteins with similar amino acid sequence, that is, between proteins closely related on the evolutionary scale. It may also occur, due to convergent evolutionary development, between proteins with no obvious sequence similarity. Therefore, the 3D structure of the query protein can be identified either on the basis of sequence similarity or, failing that, on the propensity of the query protein to adopt a particular fold. Thus, a model for the query protein can be generated by mapping the protein sequence into Cartesian space using known 3D structures of proteins as references or 'templates'. It is worth noting that analogy-based modelling does not have as high a success rate as homology-based modelling in terms of accuracy and quality of predicted structures.

Another approach to modelling protein structure is to attempt to predict the way a protein folds *ab initio*, that is, without directly using other protein structures as references.

However, this is a very unreliable method of predicting the 3D structures of proteins at atomic resolution.

Once protein structures have been modelled to atomic resolution, and the models validated, further studies may be undertaken; for example, electrostatic surface potential calculations, protein domain interaction and/or ligand binding analysis.

1.4 Protein Structural Studies Presented in this Thesis

The approaches presented in this thesis include (1) 3D structure prediction of proteins from their sequences on the basis of sequence similarity with proteins closely related on the evolutionary scale that have known structure, (2) 3D structure prediction based on the likelihood for the protein sequence to take up a particular known fold, (3) 3D structure determination using distance geometry calculations of NMR-derived distance restraint data and structural refinement, (4) assessment of generated 3D structures, (5) protein-ligand complex modelling, (6) electrostatic surface potential studies for protein interaction studies, and (7) inter-domain docking calculations.

These approaches were used, as appropriate, in the following studies: modelling the structure of human NADPH-cytochrome P_{450} reductase (Chapter 3), submission of models for the homology modelling category of the CASP2 assessment (Martin *et al.*, 1997; Chapter 4), determination of 3D structure of Cdc42Hs using NMR-derived data (Feltham *et al.*, 1997; Chapter 5), docking of a ligand into the crystal structure of a PDZ domain (Cabral *et al.*, 1996; Chapter 6), and structure/function studies of the crystal structure of trimethylamine dehydrogenase (Ertughrul *et al.*, in press; Chapter 7).

The rationale for the use of a particular modelling approach is presented in the flowchart in Figure 1.1. Several computational methods were used for more than one of the studies presented in this thesis. Therefore, all of the methods used are presented together in a single chapter (Chapter 2).

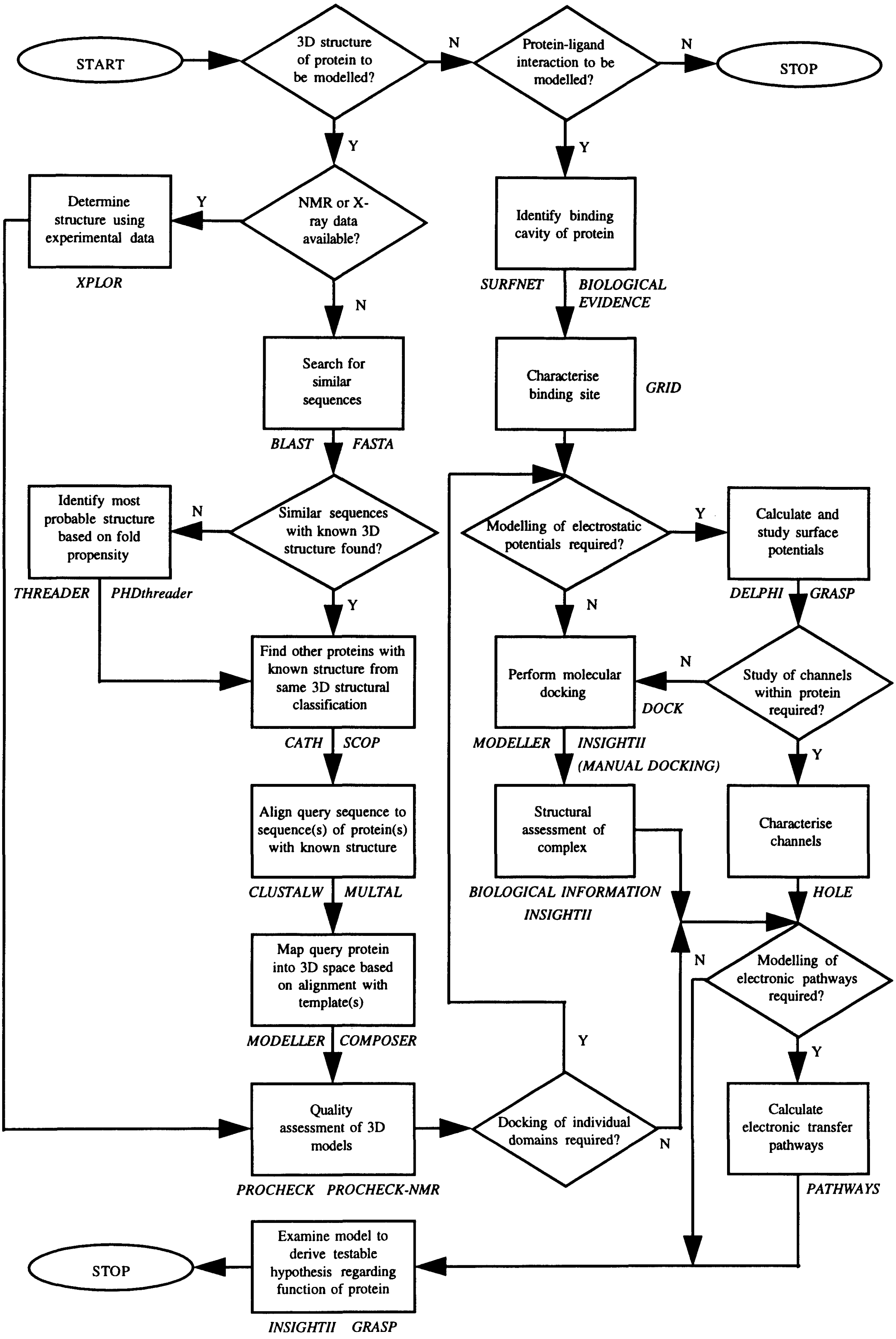

START
3D structure of protein to be modelled?
N
Protein-ligand interaction to be modelled?
N
STOP
Y
Y
Determine structure using experimental data
Y
NMR or X-ray data available?
Identify binding cavity of protein
XPLOR
SURFNET
BIOLOGICAL EVIDENCE
N
Search for similar sequences
Characterise binding site
GRID
BLAST
FASTA
Identify most probable structure based on fold propensity
N
Similar sequences with known 3D structure found?
Modelling of electrostatic potentials required?
Y
Calculate and study surface potentials
THREADER
PHDthreader
Y
N
DELPHI
GRASP
Find other proteins with known structure from same 3D structural classification
Perform molecular docking
N
Study of channels within protein required?
DOCK
CATH
SCOP
MODELLER
INSIGHTII
(MANUAL DOCKING)
Y
Align query sequence to sequence(s) of protein(s) with known structure
Structural assessment of complex
Characterise channels
CLUSTALW
MULTAL
BIOLOGICAL INFORMATION
INSIGHTII
HOLE
Map query protein into 3D space based on alignment with template(s)
N
Modelling of electronic pathways required?
MODELLER
COMPOSER
Y
Y
Quality assessment of 3D models
Docking of individual domains required?
N
Calculate electronic transfer pathways
PROCHECK
PROCHECK-NMR
PATHWAYS
STOP
Examine model to derive testable hypothesis regarding function of protein
INSIGHTII
GRASP

Figure 1.1 Flowchart of Methods Used

CHAPTER 2

COMPUTATIONAL METHODS

This chapter describes computational methods that have been applied in studying the structures of proteins presented in this thesis. Unless otherwise stated, the calculations for each computational method were performed locally on Silicon Graphics UNIX workstations with R4000 or R8000 series CPUs. The software package InsightII (MSI, San Diego, USA) was used for the graphical display and manual examination of the protein 3D structures. The software package GRASP (Nicholls *et al.*, 1991) was used for the graphical display and manual examination of surfaces, electrostatic surface potentials and their associated properties in conjunction with the 3D protein structure.

2.1 TEMPLATE-BASED MODELLING OF STRUCTURES

Modelling based on template structures of proteins can be broadly described in terms of four main steps. Firstly, suitable proteins are identified for use as templates, or structural references, for the 3D model of the query protein. The identification of template proteins refers to identification by homology (*i.e.* sequence similarity), as well as identification by analogy (*i.e.* fold recognition, also known as threading, which is based on the propensity of a protein sequence to take up a particular fold, or topology, regardless of sequence similarity).

The second step is to align the amino acid sequence of the query protein with the sequence(s) of the template protein(s). In the case of the homology-based approach, the alignment will be more accurate if it contains a larger number of proteins with sequences that are homologous to that of the query and template proteins (*i.e.* a multiple sequence alignment). In the case of the analogy-based approach, the subsequent modelling should be more accurate if the amino acid sequence of the identified template protein can be aligned with other proteins with known structure that are homologous to it (*i.e.* possess same fold) so that they be used as additional templates. It is important, in the latter case, to conserve the positional propensity of the query protein by not significantly perturbing its pairwise alignment with the identified template protein (*i.e.* to preserve the integrity of the alignment between the query protein and its identified template protein).

Thirdly, a model for the query protein is generated by mapping its sequence into 3D space on the basis of the sequence alignment, using the coordinates of the protein(s) with known 3D structure as a reference; this step also requires the structural optimisation of the query protein. Finally, the models are assessed for quality of structure.

The first three steps are discussed in the following sub-sections; the fourth is discussed in section 2.3.

2.1.1 Identification of Template Proteins

A variety of methods that identify proteins to be used as template structures for modelling the 3D structure of a query protein are described below.

2.1.1.1 Sequence Similarity

Modelling 3D structures on the basis of sequence similarity is possible because a small mutation in protein sequence usually corresponds to only small changes in protein 3D structure (Lesk & Chothia, 1986). For example, for proteins with 50% or greater sequence identity, 90% or more of the structures will be in 'core' regions (*i.e.* regions of structure in the proteins where the fold is conserved), packed secondary structural elements shift in the range of only 0.3-1.5Å with respect to each other, and over 85% of the mainchain and sidechain dihedral angle values for unmutated residues are conserved to within 30° (Lesk & Chothia, 1986).

The identification of template proteins by sequence similarity is performed by sequence database scanning using the query sequence as a reference. In order to measure homology, the comparison of sequences require quantification by the use of a scoring matrix based on amino-acid similarity. A brief introduction to some frequently-used scoring matrices, and a description of two sequence database search methods are presented below.

2.1.1.1.1 Scoring Matrices

Scoring matrices are used to measure sequence homology in a database search, and also to optimise the alignment of pairwise and multiple sequences (see section 2.1.2).

One such example of a scoring matrix is the PAM (percentage of acceptable point mutations per 10^8 years) matrix derived from the study of evolutionary mutation of residues by Dayhoff *et al.* (1978), which represents the probability of mutations (based on observed frequencies) between residue types, thereby giving a measure of relatedness between residues over a given evolutionary distance.

PAM probability matrices for a range of evolutionary distances may be generated by multiplying the original PAM matrix repeatedly by itself an appropriate number of times, after which the substitution scores may be determined. For example, the PAM40 scoring matrix corresponds to lower evolutionary distance and may be used in a database search to find regions of sequences with high similarity (thus focusing on selectivity of alignment over shorter lengths), whereas PAM250 corresponds to a higher evolutionary distance and may be used in a database search to find more general similarity (thus focusing on sensitivity of alignment over the full sequence).

The PAM matrix had its mutation frequencies derived from alignments of very similar sequences. However, an updated version of the PAM matrices, PET91, was derived by Jones *et al.* (1992a) using a larger number of sequences and protein families as a data set.

Another frequently-used set of scoring matrices is BLOSUM (blocks substitution matrix) derived by Henikoff & Henikoff (1992). The BLOSUM set of matrices, for which the data set contained more distantly related sequences than that for the PAM matrix, is derived from segments of ungapped sequence alignments, or blocks. Substitution frequencies were determined by counting matches (*i.e.* occurrences of same residue type) and mismatches (*i.e.* occurrences of different residue type) for every residue position and between every sequence in a pairwise fashion within each block, and the frequencies for each block contributed to the overall count.

However, in order to reduce ‘redundant’ counting within each block due to the existence of very similar sequences, the protein sequences were clustered into groups of similarity based on a particular percentage identity cutoff value. Thus, the counts for each match/mismatch relating to a particular sequence were averaged with respect to the counts for the same match/mismatch in the other sequences within the same group.

The BLOSUM62 scoring matrix, for example, was derived using a clustering cutoff of 62% percentage identity. A higher percentage identity clustering cutoff corresponds to a higher selectivity for residue similarity over shorter regions of sequences, whereas a lower clustering cutoff corresponds to greater sensitivity for overall similarity in more distantly related sequences.

Scoring matrices tend to be represented as the logarithm of probabilities matrices, that is, each element in the matrix is expressed as a logarithm in order to calculate the relatedness between sequences by addition rather than multiplication of probabilities.

The selection of a scoring matrix is usually a trade-off between higher selectivity over shorter regions of sequence and higher sensitivity over longer regions of sequence (*e.g.* PAM120, BLOSUM62).

2.1.1.1.2 BLAST

BLAST[1] (Basic Local Alignment Search Tool; Altschul *et al.*, 1990) is a heuristic method that searches a database of sequences in order to find those which are homologous to a

[1] URL http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-blast?Jform=1

given query sequence. The BLAST sequence search method allows for point mutations but not insertions/deletions

The BLAST method seeks to identify the 'maximal segment pair' (MSP), that is, the highest scoring segment of equal length, between the query sequence and a database sequence. The score is defined to be maximal if it cannot be improved by either extending or shortening both segments. In a database of thousands of sequences, only a few are likely to be homologous to the query sequence, and this may be considered to correspond to some MSP cutoff score, *S*.

The speed of the database search may be maximised if the time spent on regions which have little chance of exceeding such a cutoff score is minimised. A segment of fixed length, *w*, is used as a starting size (or window) for comparisons at each possible position between the query sequence and a database sequence. A fixed threshold value, *T*, is used to either accept or reject the initial comparison score. A smaller value of *T*, for example, corresponds to a smaller chance that the initial segment pair (length, *w*) will score greater than the cutoff score, *S*. If the initial segment pair score is accepted, the boundary lengths are then adjusted in order to determine the locally maximal score. A boundary extension in a particular direction is terminated where the score falls a certain distance below the best score found for shorter extensions in that direction. In practice, Altschul *et al.* (1990) found that a good choice for *w* is 4, and for *T* is 17.

The ranked results are given in terms of MSP score and the statistical significance of the MSP score, *p(N)*. The latter is given by the probability of finding a random sequence (*i.e.* a randomly shuffled version of the query sequence) of equal length to the database protein

MSP, and with a score greater than or equal to the MSP score. Also given are the ungapped pairwise segment alignments.

2.1.1.1.3 BLAST2

BLAST2[2,3] (Altschul *et al.*, 1997) is a faster version of BLAST and also allows for insertions/deletions in the alignment between query and database sequences.

As with BLAST, in BLAST2 a segment of fixed length, W (w in BLAST), is used as a starting size for comparisons between the query and database sequences, and a fixed threshold value, T, is used to either accept or reject the initial comparison score.

For the purpose of increased speed, the criterion for the extension of homologous subsequences has been modified. This is based on the observation that a BLAST 'high-scoring segment pair' (HSP; that is, a high scoring segment of equal length between the query sequence and a database sequence) of any significance is much longer than an unextended homologous subsequence, and may therefore contain more than one homologous subsequence within it.

The BLAST2, or 'gapped BLAST', method invokes an extension (in each direction) when two non-overlapping homologous subsequences are relatively close to each other, the proximity being defined by a distance parameter, A. A gapped extension is invoked for any HSP that exceeds a moderate cutoff score, S_g, which is chosen such that no more than approximately one extension is invoked per 50 homologous subsequences.

[2] URL http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-newblast?Jform=1

[3] BLAST2 became available subsequent to completion of the work presented in this thesis.

Since two homologous subsequence ‘hits’, rather than one, are required to invoke an extension, the threshold score value, *T*, used to define a ‘hit’ must be lowered to retain comparable sensitivity. The effect of lowering the threshold score is that more ‘hits’ are found; however, only a small fraction have an associated second hit in close proximity to invoke an extension.

BLAST often detects a number of ungapped alignments involving a single database sequence which become significant when taken together. By introducing a gapped alignment method, it becomes necessary to find only one gapped alignment rather than all the ungapped alignments that comprise a significant result. This is another factor that causes an increase in speed.

The default values for the above-mentioned parameters for BLAST2 are *W*=3, *T*=11 and *A*=40.

The ranked results are given in terms of the MSP score and the statistical significance of the MSP score, *E-value* (‘Expect value’; similar to the *p(N)* value in BLAST). Also given are the gapped pairwise segment alignments.

BLAST2 runs on average three times faster than BLAST, and in most cases finds a greater number of significant alignments (Altschul *et al.*, 1997). BLAST2 runs over 100 times faster than the Smith & Waterman (1981) algorithm, and seems to find nearly all the significant similarities found by this more rigorous algorithm (Altschul *et al.*, 1997).

2.1.1.1.4 Ψ-BLAST

Weak homologous relationships are better detected using a position-specific score matrix for a database search (also known as a motif- or profile-based search) than a simple sequence-based search. The motif-based search has a lower level of 'random noise' in sequence comparison than does the sequence-based search. The position-specific score matrix forms the basis of the iterative search approach, Position-Specific Iterated BLAST (PSI-BLAST or Ψ-BLAST[4,5]; Altschul *et al.*, 1997).

The position-specific score matrix describes the estimated probability for a given residue-type at each query sequence position. It is initially generated at the initial Ψ-BLAST step which is a BLAST2 query sequence search (the iterations beginning thereafter are motif-based searches). The position-specific score matrix is derived from a combination of (1) observed residue frequencies at each query sequence position in the output alignment for 'significant hits' (defined by *E-value*, defaulted to a value of 0.01), and (2) knowledge derived from the substitution matrices associated with more rigorous sequence alignment algorithms. When deriving the estimated probabilities, the aligned sequences are weighted so as to avoid biasing towards large subsets of closely-related sequences within the output alignment. The output alignment may be seen to represent a family of proteins of which the query sequence is a member; thus, a motif-based description of the alignment is also a description of the query sequence.

The first Ψ-BLAST iteration involves using the gapped BLAST methodology applied to the motif-based database search rather than the sequence-based search. That is, the estimated

[4] URL http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-psi_blast

[5] Ψ-BLAST became available subsequent to completion of the work presented in this thesis.

probabilities at each query sequence position are used as the only guide when searching the database for sequences similar to the query sequence. The position-specific score matrix is then updated based on the new search output alignment. The motif-based database search and the subsequent updating of the position-specific score matrix may then be repeated iteratively. Unless the search converges, the iteration should be performed 15 to 20 times.

The ranked results for each iteration are given in terms of MSP score and the statistical significance of the MSP score, *E-value* ('Expect value', as in BLAST2). Also given are the gapped pairwise segment alignments.

Considering only a single iteration step, the Ψ-BLAST runs faster than BLAST, and 40 times faster than the Smith & Waterman (1981) algorithm (Altschul *et al.*, 1997). Considering a search that includes only a single iteration, Ψ-BLAST finds every 'true positive' found by the Smith & Waterman (1981) algorithm but it finds many others as well, although this improves with further iterations (Altschul *et al.*, 1997).

2.1.1.1.5 FASTA

FASTA (Pearson & Lipman, 1988; also Pearson, 1990) is another database sequence search method based on gapped pairwise alignment against a query sequence. A new version, FASTA3[6] (see Pearson & Lipman, 1988; also Pearson, 1990), founded on the principles of FASTA, is now available.

[6] Available at URL http://www2.ebi.ac.uk/fasta3/

The program first matches consecutive identities of length *ktup* (*i.e.* k-tuplet, default=2 for protein sequence search) between a database sequence and the query sequence. Rather than comparing every bi-tuplet in the query sequence with every bi-tuplet in a database protein in order to test for identity, a more efficient approach is used. A look-up table for every possible amino-acid bi-tuplet is set up (*i.e.* a table of 23^2, or 529, elements), so that for each database protein the position of a particular bi-tuplet in its sequence is pre-stored under the appropriate bi-tuplet in the look-up table. Therefore, for every existing bi-tuplet in the query sequence, a (list of) match(es) can be located immediately by accessing that particular bi-tuplet in the look-up table. The ten highest pairwise regions of identity are retained in order to undergo a more rigorous sequence similarity assessment.

The ten selected regions in the protein sequence search are then scored, usually using the PAM250 matrix (Dayhoff *et al.*, 1978; see section 2.1.1.1.1). From these ten segments, those that are unlikely to constitute contiguous parts of the pairwise alignment between the query and database sequences are discarded; that is, those segments which score less than a 'joining threshold' are eliminated. The joining threshold is an empirically determined similarity cutoff score (taking into account the query sequence length and *ktup* value), and defined as being one standard deviation above the average alignment score expected from unrelated sequences in the database. Thus, the remaining segments constitute an approximate alignment.

This approximate alignment is optimised, including the joining of the segments using the gap regions, by a modification of the Needleman & Wunsch (1970) and Smith & Waterman (1981) pairwise sequence alignment algorithms[7]. The optimisation takes into account

[7] Both these algorithms compare the alignment between two sequences using a two-dimensional matrix, the elements of which contain information about (1) the similarity scores between residues, and therefore (2) the path through the matrix that corresponds to the optimal alignment positions of the residues. The

insertions and deletions, which results in increased sensitivity of the approach, although this also results in decreased selectivity for local pairwise similarity. However, the optimisation restricts residue positioning to within a narrow band based on the approximate alignment derived from the database search.

The results are ranked according to *initn*, that is, the summed score of the joined segments in the approximate alignment minus the joining threshold for each gap. Also included in the results are a significance score with respect to random shuffling of the database sequence (*z-score*), the *E-value* (similar to the *E-value* in BLAST2 and Ψ-BLAST) and the optimised pairwise alignments.

2.1.1.2 Topological Propensity

Proteins with the same fold need not be related, but could share the same fold because of the rules that govern topology and secondary structure packing (Chothia, 1992). This is shown to be true in the PDB where a number of proteins possess low sequence identity (<25%) but exhibit highly similar topology. Also, since the number of naturally occurring proteins is finite then the number of naturally occurring folds must also be finite. Chothia (1992), for example, estimates the total number of folds to be 1000, whereas Wang (1996) estimates the number to be 455. Therefore, a knowledge-base of protein structures could be used for fold recognition which is not based on sequence homology.

Needleman & Wunsch (1970) approach is optmised for global alignment of the sequences, whereas the Smith & Waterman (1981) approach is optimised for alignment based on the identification of independent regions of similar subsequences.

2.1.1.2.1 THREADER

THREADER (Jones *et al.*, 1992b) is a program that is used to identify the most probable fold of a query protein by fitting its sequence onto known folds in 3D space. The database of known folds was set up using protein structures from the PDB, with less than 30% sequence identity and with resolution of 2.8Å or higher. The fitting of the query sequence onto a fold is not based on sequence similarity but on an optimisation procedure that minimises the potential energy of the query protein.

Classical potentials cannot identify proteins that have been folded into non-native conformations, whereas empirically-derived potentials that include the consideration of residue solvation could serve to identify misfolded proteins. In this method, empirically observed frequencies of distance between backbone as well as Cβ atoms for residues over three ranges of sequence separation (*i.e.* short, medium and long range) are used to determine pairwise potentials; the THREADER method focuses on backbone conformation, and thus the remainder of the sidechains are not specifically taken into account. Also, empirically observed frequencies of percentage residue solvent accessibility are used to determine solvation potentials. The equations for the potentials are similar to those derived by Sippl (1990) from the Boltzmann Law[8].

For the pairwise potentials, the three ranges of sequence separation, k, are defined as short-range ($k \leq 10$), which predominates in the matching of secondary structural elements, medium-range ($11 \leq k \leq 30$), which mediate super-secondary structural motifs, and long-

[8] The Boltzmann Law states that a particular state, x, of a physical system in equilibrium is occupied with probability $f(x)$, which is proportional to the Boltzmann factor of state x. That is, $f(x) \propto exp[-E(x)/RT]$. Therefore, the energy of the system is given by $E(x) \propto -RT \ln[f(x)]$.

range (k > 30), which mediate tertiary packing. For the medium and long ranges of sequence separation, consideration of pairwise interactions above a spatial distance of 10Å were excluded. Since protein cores are conserved by secondary structure packing, and loop regions tend to be exposed to solvent, the pseudo-energies of loop residues are evaluated by the solvation potential only.

The expression for the pairwise potential between residues a and b is given by:

$$\Delta E_k^{ab} \quad = \quad RT\ln[1+m_{ab}\sigma] - RT\ln\left[1+m_{ab}\sigma\frac{f_k^{ab}(s)}{f_k(s)}\right]$$

where m_{ab} is the number of residue pairs ab observed at topological level k, s is the weight given to each observation, $f^{ab}{}_k(s)$ is frequency of occurrence of residue pair ab at topological level k and sequence separation distance s, and $f_k(s)$ is the equivalent frequency of occurrence of all residue pairs.

The expression for the solvation potential for residue a (calculated for all residues in the query protein) is given by:

$$\Delta E_{solv}^{a}(r) \quad = \quad -RT\ln\left[\frac{f^a(r)}{f(r)}\right]$$

where r is the percent residue accessibility (relative to residue accessibility in the GGXGG fully extended pentapeptide), $f^a(r)$ is the frequency of occurrence of residue a with accessibility r, and $f(r)$ is the frequency of occurrence of all residues with accessibility r.

The threading of the query sequence onto a particular topology is performed by optimal positioning of the query sequence residues with respect to the residue positions of the fold.

Insertions and deletions are allowed within loop regions only. The optimal alignment is defined by the lowest energy of the system.

The results are most usefully ranked with respect to two values: (1) the significance score of the pairwise energies filtered for the set of query sequences with a reasonable number of residues matched to fold positions, and (2) the significance score of the weighted sum of pairwise and solvation energy allowing for the proportion of query sequence residues matched to fold positions.

THREADER2[9] is the latest development of THREADER and contains some new features. These include (1) new statistical tests for prediction reliability, (2) the option for user-input of secondary structural elements if known or if reasonable predictions are acquired, and (3) fold library options for domain recognition.

2.1.1.2.2 TOPITS / PHDthreader

TOPITS[10], or 'Threading One-dimensional Predictions Into Three-dimensional Space' (Rost, 1995a; Rost, 1995b), also known as PHDthreader, is a neural network approach to fold recognition.

The structural data set consisted of PDB proteins with pairwise sequence identity below 30%. The DSSP[11] database (Dictionary of Secondary Structure of Proteins; Kabsch & Sander, 1983) was referenced in order to express the 3D structures of these proteins as 1D strings of secondary structure and solvent accessibility. For protein secondary structure,

[9] URL http://globin.bio.warwick.ac.uk/~jones/threader.html

[10] URL http://www.embl-heidelberg.de/predictprotein/phd_pred.html

[11] URL http://www.sander.embl-heidelberg.de/dssp/

three states were assigned: helix (H), strand (E) and loop (L). For the residue solvent accessibility, two states were assigned: buried (B, <16% solvent exposed) and exposed/outside (O, ≥16% solvent exposed). Thus, residues are expressed in terms of six structural states: H_B, H_O, E_B, E_O, L_B and L_O.

For a query protein sequence, neural network systems (trained on the data set of proteins) were used to predict the secondary structure (neural network system PHDsec; for further details, see section 2.1.2.3.1.1) and residue solvent accessibilities (neural network system PHDacc[12], Rost & Sander, 1994b; similar approach to PHDsec). The query protein is thus represented as a 1D string consisting of six structural states (see previous paragraph).

Then, the predicted 1D string of the query protein is aligned in a pairwise fashion with the projected 1D string of each database protein using a dynamic sequence alignment approach. Alignments in which the database protein is much larger than the pairwise alignment itself are ignored, resulting in an increase in prediction success. The scoring matrix used for comparing the six states of the 1D strings was derived by a trial and error method with the objective of distinguishing true homologues (*i.e.* 'correct hits', using a data set of pairwise structurally aligned remote homologues[13]) from false positives (*i.e.* 'incorrect hits', using random pairwise alignments). The gap penalty parameters (*i.e.* gap opening penalty and gap elongation penalty) were optimised with respect to given alignment lists.

[12] PHDacc predicts two-state solvent accessibility at 75% accuracy (Rost, 1995b). Two-state residue solvent accessibility is largely conserved (65%) between remote (<25% identity) homologues (Rost, 1995b).

[13] Taken from FSSP (Families of Structurally Similar Proteins) database of structurally aligned protein fold families (Holm & Sander, 1994). URL http://www2.ebi.ac.uk/dali/fssp/

The results are ranked with respect to the pairwise alignment score between the query protein and the database proteins. A significance score, *ZALI*, is also calculated for each ranked 'hit':

$$ZALI_i = (E_i - E_{mean})/\sigma$$

where E_i is the alignment score for a particular protein, E_{mean} is the average score over all hits, and σ is the standard deviation of the distribution of all hits. *ZALI* is used as the criterion for distinguishing between correct hits and false positives.

The output also includes the pairwise alignments, displaying amino acid sequence, secondary structural state and solvent accessibility state for each residue in the query and 'hit' proteins. The success rate of the method with respect to the top ranked hit is approximately 30%, increasing with higher significance score.

PHDthreader produced the best performance in the CASP2 assessment (see Chapter 4) threading category (Marchler-Bauer *et al.*, 1997).

2.1.1.3 Protein Structure Classification Resources

CATH (Orengo *et al.*, 1997) and SCOP (Murzin *et al.*, 1995) are resources that contain hierarchical classifications of proteins with known 3D structure. These resources can, after sequence similarity searches and/or topological propensity calculations, serve to ensure that proteins related to the identified template proteins are identified, thus securing a greater information base for modelling the structure of a query protein.

2.1.1.3.1 CATH

CATH[14] (Orengo *et al.*, 1997) classifies proteins hierarchically into Class, Architecture, Topology and Homologous superfamily. The maintenance and updating of CATH is largely automated by the use of suitable algorithms for protein structure classification. Approximately 90% of the structures can be classified automatically, the remainder requiring manual classification (Michie *et al.*, 1996). The CATH resource also provides additional information such as a PROCHECK (Laskowski *et al.*, 1993; see section 2.3.1) summary for quality of structure and per-residue secondary structure assignment.

For 'Class', the proteins are classified slightly differently to SCOP, falling into α, that is, proteins of which the structures are formed essentially by α-helices, β, that is, proteins of which the structures are formed essentially by β-sheets, αβ (which can be further subdivided into α/β and α+β categories, where α/β refers to proteins with largely interspersed α-helices and β-strands, and α+β refers to proteins with largely segregated α-helices and β-strands) and 'Few Secondary Structures'. 'Architecture' describes the general secondary structure arrangement without referring to topological details; the assignment for protein architecture is currently performed manually. 'Topology' details the overall shape and secondary structure connectivity, and is specified by reference to domain type. 'Homologous superfamily' groups proteins which can be considered to share a common evolutionary ancestor based on sequence and structural similarity, and is defined by the PDB accession code of a representative structure.

[14] URL http://www.biochem.ucl.ac.uk/bsm/cath/

For example, the *Desulfovibrio vulgaris* flavodoxin (using PDB accession code 2fx2) is classified in CATH as follows:

Class	αβ
Architecture	3-layer (αβα) sandwich
Topology	Nitrogen Regulatory Protein NTRC Receiver Domain
Homologous superfamily	4fxn
Protein	*2fx2*

2.1.1.3.2 SCOP

SCOP[15] (Murzin *et al.*, 1995), which is maintained and updated manually, classifies proteins hierarchically into Class, Fold, Superfamily, Family and Domain. 'Class' classifies proteins as follows: α, β, α/β and α+β. 'Fold' refers to packing and topology of secondary structure in 3D space. 'Superfamily' relates to proteins with low sequence identities but the structures (and often the functional features) of which indicate a probable common evolutionary origin. 'Family' relates either to proteins with sequence identities of 30% or more, or with lower sequence identities but for which the 3D structures and functions are very similar. 'Domain' refers to distinct structural and/or functional subsets within a protein.

For example, the *Desulfovibrio vulgaris* flavodoxin (using PDB accession code 2fx2) is classified in SCOP as follows:

Class	α/β
Fold	Flavodoxin-like 3 (α/β/α) layers; parallel β-sheet of 5 strands, order 21345
Superfamily	Flavoprotein
Family	Flavodoxin
Protein	*2fx2 (consists of a single domain)*

[15] URL http://scop.mrc-lmb.cam.ac.uk/scop/

2.1.2 Multiple Sequence Alignment

Once sequences that are similar to the query sequence have been identified, it is necessary to align them into a single body of multiple sequences in order to represent their relationship with respect to each other.

2.1.2.1 Automated Alignment of Multiple Sequences

Multiple sequence alignments can be considered to fall into two broad approaches: the progressive and the single order approach. In both cases, the individual sequences are first aligned in a pairwise fashion and scored on sequence similarity.

Then, in the progressive approach, the multiple alignment begins by aligning the two sequences with highest similarity score. Following this, the next highest similarity score is then considered, which may lead to the alignment of two other sequences, or may require the incorporation of a sequence into the already existing alignment. This continues until all the sequences are incorporated into the final alignment.

The positional relationship within already-existing sub-alignments is preserved, thus not compromising the alignment between more closely related sequences by more distantly related ones. For the alignment between sub-alignments, the score at a particular residue position is the average of the individual pairwise residue similarity scores.

In contrast, for the single order approach, no sub-alignments (*i.e.* intermediary alignments) are generated, but each single sequence is incorporated in turn into one multiple alignment, based on that sequence having the highest similarity score with any sequence already in the alignment.

There is little advantage in choosing one alignment approach over the other when aligning very similar sequences. However, the progressive approach is more suited to aligning less similar sequences, because it is conducive to the clustering of sequence sub-families within the alignment.

2.1.2.2.1 CLUSTALW

CLUSTALW (Thompson *et al.*, 1994) is the latest development of the CLUSTAL approach (Higgins & Sharp, 1988).

The CLUSTALW alignment may be performed using a choice of either a 'fast' or 'accurate' pairwise alignment method. The 'accurate' method has acceptable memory requirements[16], and is therefore preferable since alignment speed is satisfactory. For this method, the alignment is generated by first matching two similar residues, or a residue and a gap, at a mid-point position of each sequence to form the first alignment pair; then optimal mid-points on either side of the first optimal mid-point are found recursively until the sequences are fully aligned with respect to each other.

The CLUSTALW method aims to take into account the wide range of similarities between protein sequences by incorporating (a) the weighting of sequences in an alignment, (b) the

[16] The memory requirement is linearly related to sequence length, whereas, for example, for the Needleman & Wunsch (1970) method the memory requirement is proportional to the square of the average sequence length.

variation of substitution matrices within the alignment process, and (c) an improved treatment of gap penalties.

In the multiple sequence alignment procedure, groups of closely related sequences contain a high amount of duplicate information, and are therefore assigned lower weightings, than more divergent sequences, which are assigned higher weightings (proportional to sequence dissimilarity), in the calculation of the sequence contribution of residues when aligning two sub-alignments. This weighting is automatic and takes into account the relationship between the sequences, and is thus more representative of them than merely using an average-based approach.

A choice of two types of substitution matrix may be used, either PAM (Dayhoff *et al.*, 1978; see section 2.1.1.1.1) or BLOSUM (Henikoff & Henikoff, 1992; see section 2.1.1.1.1). In a particular alignment step, one of four matrices of each type are used automatically (PAM 20/60/120/350 or BLOSUM 80/62/45/30) depending on the similarity between the two sequences, or the two groups of sequences, to be aligned.

Two gap penalties are used, the values of which are initially set by the user: a gap opening penalty (GOP) and a gap extension penalty (GEP). The CLUSTALW program then attempts to automatically select appropriate gap penalties for each pairwise alignment. A number of factors contribute to the automatic selection of the gap penalties. For example, a more similarly matched pair of sequences are assigned a higher GOP; if one sequence is much shorter than the other, the GEP is increased to inhibit too many long gaps in the shorter sequence.

The gap penalties are also specific to residue position and the following rules are applied hierarchically. (1) If there is a gap at a position then both the GOP and the GEP are decreased (and rules 2 to 4 are ignored); this makes gaps more likely at positions where there are already gaps. (2) If there is no gap at a position and if this position is within 8 residues of another gap then the GOP is increased; this discourages the creation of gaps that are too close together. (3) If a residue is within a segment of 5 or more hydrophilic residues then the GOP is decreased; these areas usually indicate loops regions in proteins. (4) If the segment does not contain 5 or more hydrophilic residues then the GOP is modified according to residue-specific gap propensities; the propensities were derived from frequencies of each residue at either end of gaps in alignments of proteins with known 3D structure.

2.1.2.2.2 MULTAL

MULTAL (Taylor, 1988) performs a pairwise alignment of sequences using a modified version of the Needleman & Wunsch (1970) algorithm, where CPU time is saved by applying a *window* parameter to the two-dimensional sequence comparison matrix, so that only residues within a specified range from the diagonal (*i.e.* relatively close pairwise proximity) are compared.

The scoring is defined as a relative contribution between the identity matrix (ID) and one chosen from a given selection (M) such as that of Dayhoff *et al.* (1978) or Henikoff & Henikoff (1992). The contributions are defined by the matrix weight parameter, *matwt*, ranging in value from 0 to 10. If *matwt*=8, for example, then the score is calculated as 0.8M+0.2ID. An alignment might begin with *matwt*=0 in order to align more similar sequences, then for subsequent cycles *matwt* may be increased by the user to take better account of the alignment of less similar sequences.

The gap penalty has a single value, of which the optimum for a particular set of proteins is found by trial and error.

An alignment or sub-alignment is allowed to form only if the pairwise score for the relevant (sets of) sequences is above a cutoff score. The cutoff score is set higher in earlier cycles so that only sequences with higher similarity are allowed to form sub-alignments. The cutoff is gradually decreased for successive cycles so that all the sub-alignments eventually become incorporated into a single alignment.

MULTAL can also take into account the order of sequences in a list that is arranged according to similarity using the *span* parameter. The *span* value is the number of sequences to which each sequence is compared, so that *span*=3 allows comparison of sequences up to three sequences apart in the list. If the sequences have not been ordered with respect to their similarity, *span* should be set to the number of sequences in the list.

2.1.2.3 Refinement of automatically aligned sequences

The alignment generated from an automatic sequence-based alignment method often requires manual adjustment because the sequence alignment calculation is very parameter sensitive, and also because not all the structural information is taken into account by primary sequence similarity alone. The alignment can therefore be manually refined based on the secondary and tertiary structural information of the aligned proteins. Such information can be obtained directly from proteins with known 3D structures, for which the alignment can be refined with respect to each other based on their tertiary structures.

2.1.2.3.1 Secondary Structure Prediction

A multiple sequence alignment generated using an automatic method can be refined for proteins for which the 3D structures are unknown, using its predicted secondary structure against the known secondary structures of the template proteins in the alignment. Care must be taken when using predicted secondary structural information for refinement of the sequence alignment due to the uncertainty in secondary structure assignments. Three secondary structure prediction methods are discussed below.

2.1.2.3.1.1 PHDsec

The PHD (Profile network from HeiDelberg) secondary structure prediction tool, PHDsec[17] (Rost & Sander, 1993a; Rost & Sander, 1993b; Rost & Sander, 1994a; Rost *et al.*, 1994), is based on seven independent neural networks. Six of these networks consist of two consecutive levels, that is, a 'sequence-to-structure' network level followed by a 'structure-to-structure' network level. One of the seven networks consists of a 'sequence-to-structure' network level only. For a given residue position in a protein sequence, the output from each network, and from each level within a network, specifies three values between 0 and 1, which correspond to the probabilities of the residue being in a particular secondary structural state (*i.e.* helix, strand, loop). The resulting output from each of the seven independent networks is fed into a final and common level, where a secondary structural

[17] URL http://www.embl-heidelberg.de/predictprotein/phd_pred.html

state for a residue position is assigned by 'jury decision' based on the results of the seven independent networks.

Within a network, each of the first two network levels consists of nodes (*i.e.* processing subunits in the network) which are interconnected according to a particular architecture. The output from a given set of nodes of number, j, which are to be used as input for a subsequent node, are assigned with individual strengths, or weightings, J_j. Thus, each input is 'valued' by the subsequent node according to the strength corresponding to its associated input node. The strength values were optimised through the 'training' of the seven networks, each of which utilised a particular combination of 'learning' and 'testing' proteins from 130 protein chains (with resolutions of 2.5Å or higher) from the PDB. The optimisation was terminated once the secondary structure prediction accuracies for a network's training proteins was greater than 70% for the first network level, and greater than 75% for the second network level.

To predict the secondary structure of a protein sequence, PHDsec first scans the SwissProt[18] database for similar sequences (30% or higher sequence identity) using the FASTA approach (Pearson & Lipman, 1988; also see section 2.1.1.1.5), and then aligns all of these similar sequences using the MaxHom algorithm (Sander & Schneider, unpublished; based on Smith & Waterman algorithm, 1981). The PHDsec server can also accept a query sequence that is already-aligned with similar sequences. For the 'training' procedure, alignments were taken from HSSP[19] (database of Homology-derived Structures and Sequence alignments of Proteins; Sander & Schneider, 1991).

[18] URL http://expasy.hcuge.ch/sprot/sprot-top.html

[19] URL http://www.sander.embl-heidelberg.de/hssp/

The first network level (*i.e.* sequence-to-structure network) derives the secondary structural state probabilites for each residue using a 13-residue window to extract a local amino acid profile (*i.e.* information regarding sequence composition). Four of the seven independent networks use the alignment of the query sequence with similar sequences for input, which provides greater sequence information. The output from the first level network is used as input for the second level network (*i.e.* structure-to-structure network) which derives the secondary structural state probabilites for each residue using a 17-residue window; thus, this level takes into account local secondary structure composition. The 'jury decision' is the secondary structural state with the highest probability value given by the average of the seven independent networks.

A reliability value is assigned to the finally predicted secondary structural state for a particular residue position in order to express the level of agreement between the 'jurors'. The reliability value is an integer ranging from 0 to 9 (where a value of 9 corresponds to a 'reliable' prediction).

2.1.2.3.1.2 GOR II

GOR II is incorporated into the Cameleon (Oxford Molecular Ltd., Oxford, UK) sequence alignment software package. It is founded on the principles of the GOR method (Garnier *et al.*, 1978) and uses a larger number of 3D protein structures for the data set.

The method is based on the observation that the amino acid composition is different in each secondary structure class. The probability of a residue, *i*, being in a particular state, *s* (helix, sheet, turn, or coil), depends not only on its own amino acid type but also on that of its sequential neighbours (Robson & Suzuki, 1976). The statistical analysis of backbone conformation within 3D protein structures by Robson & Suzuki (1976) shows that the conformation of a particular residue is strongly affected by other residues up to a separation of eight amino acids. Thus, in the GOR method, the preference for residue *i* to be in state *s* is calculated using a window of 17 residues (*i.e.* *m*-8 to *m*+8).

A set of tables is used for the secondary structure prediction. Each table corresponds to a particular secondary structural state, *s*, and contains 'preference values' for every amino acid type in each position of the window. The preference values of Garnier *et al.* (1978) are the 'information measure' values of Robson & Suzuki (1976) which relate to the probability of the existence of a given conformational state due to particular residue types at varying sequence separations.

To predict the secondary structure of residue *i*, the preference values are summed over residues *m*-8 to *m*+8 for each state, *s*, to give score, I_s. That is,

$$I_s = \sum_{j=i-8}^{j=i+8} weight_j$$

The score, I_s, is modified for each secondary structure state by a decision constant, D_s, to give the final score, $I_{s,final}$, for that state:

$$I_{s,final} = I_s - D_s$$

The decision constant is an adjustable parameter that defaults to a value relating to the ratios of states of residues found in 3D protein structures. The predicted secondary structural state for a residue in a given protein sequence is defined by the state with the highest final score.

The GOR/GORII method is not as reliable as the PHDsec method.

2.1.3 Generation of 3D Models

3D model generation of the query protein is performed after the alignment of the query sequence with related sequences, at least one of which must have known 3D structure. Instances of two different approaches are described here. Firstly, modelling of 3D structure based on the satisfaction of spatial restraints, where all the components of a protein structure are modelled simultaneously, and secondly, the fragment-based approach, where the protein is broken down into distinct parts comprised of structurally conserved and structurally variable regions. The 'satisfaction of spatial restraints' approach is the only method which allows (1) ligands / co-factors to be incorporated into the modelling building of the protein structure, and (2) the incorporation of user-defined distance restraints within the modelling calculations.

2.1.3.1 MODELLER

The 3D structure predictions of proteins in this thesis have been performed using MODELLER (Sali & Blundell, 1993), a set of programs that generates models of proteins based on the satisfaction of spatial restraints that are used to describe protein 3D structure. This section details the MODELLER approach and its underlying principles.

2.1.3.1.1 Description of Structural Features in Proteins using Spatial Restraints

Instances of particular structural features can be assigned to a query protein based on the existence of such features in template proteins (*i.e.* proteins in the sequence alignment with experimentally determined 3D structure).

Considering, for example, backbone inter-residue $N_{(H)} \cdots O=C$ hydrogen bonds, if the backbone heavy atoms of two particular residue positions have suitable hydrogen bonding geometry in all the template structures, that is, if they are within hydrogen bonding distance and possess suitable angular positioning, then it can be assumed that the backbone heavy atoms of the equivalent residues in the query protein will also form hydrogen bonds. Therefore, the distance between these atoms in the query protein is appropriately restrained when attempting to generate a 3D model for the query protein.

The value of each instance of a structural feature in the query protein can be represented by an empirically-derived probability density function (PDF). A PDF is expressed in the form of a continuous function, *p(x)*, that describes the probability of occurrence with respect to a given value, *x*, for a particular structural feature. The function *p(x)* is non-negative and integrates to unity over all possible values of *x*.

The query protein, as a whole, can be described by taking into account all of its individual PDFs, the molecular PDF being the product of its component PDFs.

$$P_{mol} = p(x_1).p(x_2).p(x_3).p(x_4)\ldots p(x_n)$$

A representative selection (Sali & Blundell, 1993) of the PDB was used as a data source for the derivation of PDFs. The resolution for the structures ranged from 1.5Å to 2.9Å, only a small number of which were of lower resolution than 2.5Å.

The PDFs were derived by first determining the frequency of occurrence of every value for each structural feature in a particular topological position in a protein. The frequencies were normalised with respect to the total number of occurrences for a particular structural feature to give relative frequencies. Then, an analytical function, the PDF or *p(x)*, was fitted to the observed frequency data using a least squares algorithm. Thus, the use of a PDF is a more complete way of describing the value of a particular structural feature than, for example, merely using the mean of empirical values or the use of upper and lower bounds only. However, the PDF is only an approximation to the 'true' PDF due to the use of a non-exhaustive set of protein structures for the extraction of the frequency data.

2.1.3.1.2 Instances of Structural Features in Terms of Spatial Restraints

PDFs are used by MODELLER to describe the 3D structure of a protein and incorporate a number of structural features including mainchain conformation, sidechain conformation, inter-residue N···O distance, Cα···Cα distance, and a number of stereochemical features. The set of stereochemical features accounted for in MODELLER consists of bond lengths, bond angles, planarity of peptide groups, planarity of sidechain rings, chirality of Cα atoms, chirality of sidechain atoms, van der Waals contact distances and also the bond lengths, bond angles and dihedrals angles of disulphide bridges. Distance restraints, such as those derived from NMR spectroscopy, and/or those derived from biological knowledge (*e.g.* H-bonds, disulphide bridges) may also be specified within MODELLER, which allows the modelling of ligands and prosthetic groups into a protein. The PDFs for some of the structural features incorporated into MODELLER are described in more detail below.

Mainchain conformation is expressed using PDFs for both ϕ and ψ dihedral angles. Although two PDFs are used to define mainchain conformation, the values of ϕ and ψ dihedral angles are highly correlated, as conveyed by the relatively distinct clustering of ϕ and ψ values: the distribution of ϕ and ψ values have been shown to fall into six conformational classes, each with an empirically derived mean and standard deviation for both its ϕ and ψ dihedral angles (Wilmot & Thornton, 1990). Conformational class is the primary feature in determining the PDF for mainchain conformation in MODELLER.

Continuous distribution curves in the form of Gaussian functions are fitted to the empirically derived distributions of Wilmot & Thornton (1990) for both ϕ and ψ dihedral angle values in all of the six conformational classes. However, the PDF for mainchain conformation is dependent on residue type, and must take into account the probability of each residue type being in a particular conformation class. This probability has been determined based on observed data from the PDB protein structural database. The PDFs, with respect to both ϕ and ψ dihedral angles, for a particular residue type can then be expressed as a residue-specific weighted sum of the six mainchain conformational classes (A,B,P,G,L,E; see Wilmot & Thornton, 1990), which are each described by a Gaussian functions, N, with class weighting ϖ_i, mean α_i and standard deviation σ_i. That is:

$$p_\phi = \sum_{i=A,B,P,G,L,E} \omega_i N\left[\alpha_i(\phi), \sigma_i(\phi)\right]$$

$$p_\psi = \sum_{i=A,B,P,G,L,E} \omega_i N\left[\alpha_i(\psi), \sigma_i(\psi)\right]$$

However, using a subset of the proteins with known structure as a test set it was found that, as well as using residue type to determine the PDFs for conformational class, two other factors are important predictors of the mainchain conformation. The first is the residue neighbourhood difference between the query and template proteins, which is a measure of the dissimilarity of residues in the 3D environment of a particular residue using a nearest atom contact distance of 6Å; the contact residues in the query protein are those equivalent to contact residues in the template protein. The second factor, and more significant than either residue neighbourhood difference or residue type, is the conformational class of the equivalent residue in the template protein. Both of these factors are incorporated into the weightings for the PDFs described in the above equation.

Sidechain conformation is represented in a similar way to mainchain conformation, using a weighted sum of Gaussian functions to define the PDFs for the χ_1, χ_2, χ_3 and χ_4 sidechain dihedral angles. That is:

$$p_\chi = \sum_i \omega_i N[\alpha_i(\chi), \sigma_i(\chi)]$$

The PDF for backbone inter-residue N···O distance values are represented as a function of difference between the value in the model (h) and the value in the equivalent position in the template protein (h'). It is represented in the form of a Gaussian function with a mean of zero:

$$p_{N\cdots O} = \frac{1}{\sigma_h\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}\left(\frac{h-h'}{\sigma_h}\right)^2\right]$$

The standard deviation, σ_h, for this PDF is calculated as the sum of a large number of individual terms involving four independent variables, namely (1) the specific inter-residue

N···O distance in the template structure, (2) the fractional sequence identity, that is, the number of identical residues in the pairwise sequence alignment between the query protein and the template protein divided by the length of the shorter protein sequence, (3) the average of the fractional solvent accessibilites for the two relevant residues in the template structure, where the fractional solvent accessible area for a residue is obtained by dividing the contact area of that residue in its particular 3D environment by the standard contact area of its residue type in the extended tripeptide Gly-X-Gly, and (4), for the two relevant residues in the template protein, the average number of residue positions from a gap (insertion/deletion) in the pairwise sequence alignment with the query protein sequence. The fourth parameter was selected on the basis that the 3D structure varies more for residues closer to a gap in the sequence alignment.

The PDF for Cα···Cα distance is determined in the same way as for backbone N···O distance, and is also represented as a function of difference between the value in the model (*d*) and the value in the equivalent position in the template protein (*d'*):

$$p_{C\alpha\cdots C\alpha} = \frac{1}{\sigma_d\sqrt{2\pi}}\exp\left[-\tfrac{1}{2}\left(\frac{d-d'}{\sigma_d}\right)^2\right]$$

However, the values of the coefficients for the individual terms that constitute the standard deviation, σ_d, are different to that of the N···O standard deviation.

MODELLER can also incorporate user-specified distance constraints (defining upper and/or lower bounds), such as those between a protein and its ligand, using the Gaussian treatment of a harmonic function, as above. The template structures are used to calculate the values for the mean and standard deviation for the PDF function.

The values for stereochemical features are derived from other works based on crystal structures of small molecules such as those in the Cambridge Structural Database[20], spectroscopic studies and theoretical calculations. The PDF for bond distance, for example, is determined from the classical harmonic model for the potential energy, E_b.

$$E_b = 0.5k_b(b - b_o)^2$$

where k_b is the force constant, b is the 'observed' bond length and b_o is the 'ideal' bond length. A classical statistical mechanics treatment of this energy function produces a Gaussian probability density function which is dependent on two parameters, the bond length mean and standard deviation (Hill, 1960). That is:

$$p_b = N(\bar{b}, \sigma_b) = \frac{1}{\sigma_b\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{b-\bar{b}}{\sigma_b}\right)^2\right] \quad \text{where } \sigma_b = \sqrt{k_B T / k_b} \text{ and } \bar{b} = b_o$$

where k_B is the Boltzmann constant and T is the absolute temperature.

The derivation of the bond angle PDF is equivalent to that of the bond length. Also equivalent are derivations of PDFs that have monomodal distributions for structural feature values, such as dihedral angles for ring planarity, and the chirality of Cα atoms.

2.1.3.1.3 Satisfaction of Spatial Restraints for Protein 3D Structure Generation

As a preliminary stage before 3D structure generation of the query protein, the template proteins are first spatially aligned, normally with respect to Cα atoms, using the principle of

[20] URL http://www.ccdc.cam.ac.uk/

least squares fitting and the new coordinates are written to individual PDB format files. The equivalent Cα atoms used for superposition are defined on the basis of the pre-determined multiple sequence alignment. The structurally equivalent residues in the template proteins are identified and updated automatically by MODELLER through the iterative superposition process using a cutoff distance of 3.5Å. After superposition of the template proteins, model generation for the query protein may be performed.

The framework, or starting structure, for the 3D model of the query protein (which corresponds to a starting position for the objective function before optimisation) is based on the mean coordinates of equivalent residues in the template proteins derived from the sequence alignment. The regions for which there are no equivalent positions in the template structures are currently assigned with extended mainchain conformation (a more sophisticated loop generation algorithm is now available version 4 of MODELLER), and the sidechains are built based on knowledge derived from families of proteins in the PDB. The atomic positions of the starting structure is randomised to within a user-defined distance to allow an ensemble of models to be derived for the query protein.

For the purpose of 3D structure modelling, PDFs are assigned for all the individual structural features within the query protein. The optimal 3D structure of the query protein corresponds to the most probable values for its structural features. Thus, the 3D structure of the query protein may be optimised by maximising the molecular PDF, P_{mol}. However, the optimisation is performed using the objective function, F_{obj}, where

$$F_{obj} = -\ln\left(P_{mol}\right)$$

The use of the logarithm allows the addition of terms to produce the objective function rather than the multiplication of terms to produce a molecular PDF (thus reducing CPU time and floating point errors). Introducing the negative term in the objective function allows the 3D structure of the query protein to be optimised by the minimisation of the objective function, F_{obj}, rather than by the maximisation of the molecular PDF, P_{mol}.

The optimisation utilises the variable target function approach (originally used by Braun & Go, 1985, for optimisation of dihedral angles rather than Cartesian coordinates) incorporated into conjugate gradients minimisation and simulated annealing. That is, the optimisation begins by first satisfying sequentially local restraints, and then gradually introduces longer-range restraints until all the restraints that describe the query protein structure are included. Both the variable target function and simulated annealing approaches are strategies used to avoid the optimisation process becoming trapped in a local minimum.

It is desirable to generate an ensemble of models, rather than producing only a single model, for two reasons. Firstly, to identify structurally conserved and variable regions within the model, and, secondly, to identify structural feature violations that occur consistently across the whole ensemble. The latter indicates regions in the multiple sequence alignment that may require correction after which the structure generation procedure will need to be repeated. The process of iteratively updating the sequence alignment after 3D model generation (aided by interactive molecular graphics) should be continued until consistent structural violations are removed.

Finally, the quality of the generated 3D structures is assessed using relevant validation software such as PROCHECK (Laskowski *et al.*, see section 2.3.1) and PROCHECK-NMR

(Laskowski *et al.*, 1996; see section 2.3.1). Furthermore, a single structure may be selected as a representative of the ensemble on the basis of cluster analysis using a suitable set of algorithms such as NMRCLUST (Kelley *et al.*, 1996), and core atoms defined using NMRCORE (Kelley *et al.*, 1997).

2.1.3.2 COMPOSER

COMPOSER (Sutcliffe *et al.*, 1987a,b) uses a fragment-based approach to generate the 3D model of a query protein. This method is very successful where the known structures cluster around that to be predicted, and where the sequence identity of the query sequence to the template(s) is high, that is, greater than 40% (Sali *et al.*, 1990).

The user firstly defines structurally conserved regions (SCRs) in the query and template proteins on the basis of the sequence alignment. The program performs a Cα superposition of the template protein structures with respect to the equivalent residues in the SCRs using the method of least squares. A framework of SCR fragments is then derived from the superposed template structures.

The framework is determined by first deriving its Cα positions using an iterative process to produce a sub-structure which is self-consistent. The first approximation of the Cα positions in the framework is the SCR of a randomly selected template protein. Each template is then superposed onto this first approximation in turn, and the average of the SCRs in the newly positioned templates defines the next approximation for the framework. If two consecutive frameworks have not converged (*i.e.* rms difference $>10^{-5}$Å), then the templates are superposed onto the most recent approximation of the framework in order to determine a new framework. The self-consistency of the two most recent frameworks is tested again. The framework generation is repeated until self-consistentcy is achieved.

The structurally variable regions (SVRs) are then built into the gaps in the query protein SCRs. Suitable fragments are derived by searching a library of fragments using distance constraint criteria in order to match the endpoint geometries of the fragments in the framework. From the shortlisted SVR fragments, one is selected on the basis of possessing correct sequence length and highest sequence similarity with the corresponding SVR of the query protein in key fragment positions (residue 3D spatial positions that are common among shortlisted loop fragments). The automated selection of particular SVRs may be overridden if a specific fragment of protein structure is thought by the user to be a better solution than that which is computed.

Each sidechain is built using a procedure that draws from a set of structural rules to select the equivalent sidechain in the template proteins with best substitution score (based on the correlation in spatial position between the residue type in the template protein with the residue type in the query protein in a particular topological position), and then to replace the appropriate template sidechain with that of the residue type in the query protein using a least squares fit with suitable atomic weightings. The set of rules consists of one rule for each of the 20 by 20 amino acid replacements for each of three topological positions (*i.e.* α-helix, β-strand, loop) derived from the 3D structural analysis of proteins in the PDB database.

The final model of the query protein is then energy minimised using appropriate software in order to remove minor structural inconsistencies. These inconsistencies usually refer to steric overlap of sidechains and positions in the mainchain where an SVR joins an SVR.

2.2 3D STRUCTURE GENERATION FROM NMR DATA

The determination of 3D structures of proteins from NMR spectroscopy can be defined in terms of three main steps. First, the collection and interpretation of experimental data in order to produce both distance and dihedral angle constraints. Second, the mathematical treatment of the distance restraints in order to derive a set of starting structures. Third, the optimisation of the starting structures to give a set of 'refined' structures. These steps are discussed in the sections below. For the application presented in this thesis, step one was performed by collaborators, and steps two and three were performed using XPLOR version 3.843 (Brunger, 1996).

2.2.1 NMR-derived constraints

Multi-dimensional (*e.g.* 2D, 3D) NMR spectra exhibit cross-peaks that result from interactions between nuclei with non-zero nuclear spin (*e.g.* ^{1}H, ^{13}C, ^{15}N) from particular chemical shift positions (which are defined by their location on the diagonals of the spectra).

A COSY (*i.e.* COrrelation SpectroscopY) experiment produces peaks between hydrogen atoms that are covalently connected within three bonds. The objective of a COSY experiment is to determine the network of spin-spin couplings in a molecule in order to arrive at spectral assignments (*i.e.* to resolve which chemical shift corresponds to which nucleus). This information is useful, for example, in the analysis of an NOESY (*i.e.* Nuclear Overhauser Effect SpectroscopY) spectrum.

A NOESY experiment produces peaks between pairs of hydrogen atoms (due to 'through-space' nuclear interaction) that are spatially nearby (up to 5Å) regardless of primary sequence position. The distance between two particular hydrogen atoms, r_{ij}, is determined from the intensity of its NOE peak, I_{ij}, using a reference distance, r_{kl} (*e.g.* the distance between two particular protons on a β-strand), and the corresponding reference NOE intensity in the spectrum, I_{kl}. The relationship is given by

$$r_{ij} \quad = \quad r_{kl}\left(\frac{I_{kl}}{I_{ij}}\right)^{\frac{1}{6}}$$

Distance constraints are usually specified in terms of lower bounds (determined from the sum of the van der Waals radii) and upper bounds. These are usually categorised on the basis of the NOE intensity (*e.g.* strong:1.8-2.5Å, medium:1.8-3.5Å, weak:1.8-5.0Å).

Mainchain φ angle constraints may be derived from HN-Hα vicinal coupling constants ($^3J_{HN-H\alpha}$) using the Karplus equation (Karplus, 1959), as follows:

$$^3J_{HN-H\alpha} \quad = \quad A\cos^2(\phi - 60°) - B\cos^2(\phi - 60°) + C$$

where $^3J_{HN-H\alpha}$ is in Hertz, and *A*, *B* and *C* are constants derived by fitting measured coupling constants ($^3J_{HN-H\alpha}$) to torsion angle φ from 'known' protein structures. This quadatric equation suggests ambiguous values for φ. However, φ is generally negative for all amino acids except for glycine, aspartate and asparagine, and these values are concentrated in the range -30° to -180°. χ_1 dihedral angle constraints may be derived using $^3J_{H\alpha\text{-}H\beta}$ vicinal coupling constants in conjunction with the intensities of intra-residue Hα-Hβ and HN-Hβ NOE peaks.

Protein secondary structure determination may be deduced using a combination of NOE constraint values (*e.g.* β-strands[21] are characterised by strong $H\alpha_i$-HN_{i+1} and weak HN_i-HN_{i+1} NOE peaks, where *i* is the residue number), HN-Hα coupling constants (*i.e.* $^3J_{HN\text{-}H\alpha} \leq 5Hz$ corresponds to a helical conformation and $^3J_{HN\text{-}H\alpha} \geq 8Hz$ corresponds to a β-strand conformation), amide proton exchange rates (low rates of exchange with water indicates that the amide proton is involved in a hydrogen bond) and chemical shift index (a reference to allow a secondary structural element to be assigned from chemical shift value; Wishart & Sykes, 1994; Wishart *et al.*, 1992). The dihedral angles for secondary structural elements are defined in terms of φ (*e.g.* α-helix -55±20°, parallel β-strand -120±20°, anti-parallel β-strand -140±20°) and ψ (*e.g.* α-helix -45±50°, parallel β-strand -115±50°, anti-parallel β-strand -135±50°) constraints. ψ constraints are determined empirically from PDB and not directly from the NMR experiment being undertaken.

Once secondary structural elements have been determined, hydrogen bond distance constraints may also be assigned (*e.g.* 1.8-2.3Å for hydrogen-oxygen distances, 2.5-3.3Å for nitrogen-oxygen distances) if the existence of those hydrogen bonds is considered to be veritable on the basis of amide proton exchange rates (hydrogen bonds are indicated by lower exchange rates), NOEs and chemical shift index.

[21] NOE intensities between adjacent β-strands within a β-sheet (*i.e.* NOE distances for $H\alpha_i$-$H\alpha_j$, $H\alpha_i$-HN_j and HN_i-HN_j, where *i* denotes a hydrogen atom on one β-strand and *j* denotes a hydrogen atom on the second β-strand) not only provide a further means of identifying regions of β-strand, but also serve to distinguish between parallel and anti-parallel β-strands.

Cα and Cβ secondary chemical shift data (*i.e.* the difference between observed and random coil shifts) are correlated to ϕ/ψ space in proteins, and may be used for structural refinement (Kuszewski *et al.*, 1995a). For this purpose, ϕ/ψ space can be represented by a grid of 2° by 2° squares, and each grid corresponds to 'expected', or empirically-derived (Spera & Bax, 1991), values for the Cα and Cβ secondary chemical shifts. The backbone conformation of the protein under study is described by comparing experimental and expected chemical shift values, where the secondary chemical shifts for both Cα and Cβ are taken into account.

Proton chemical shift restraints may also be included for structural refinement, where the experimental shifts are compared to expected shifts (Kuszewski *et al.*, 1995b). The expected shifts are calculated from the geometry and charge of associated atoms.

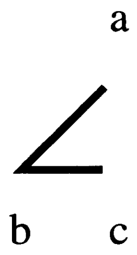
2.2.2 Generating Starting Structures: *XPLOR*

'Metric matrix distance geometry' (Havel *et al.*, 1984) is a method that is used to convert distance constraints, defined by 'ideal' bond lengths and distances corresponding to 'ideal' bond angles, as well as NOE peaks from NMR spectra, into a set of (sub-)structures represented by Cartesian coordinates. If sub-structures rather than complete structures are generated, they generally consist of a limited set of protein atom types (*e.g.* HN, N, Hα, Cα, C, Cβ and Cγ) to represent the protein; this is advantageous is the 'embedding' and optimisation stages (discussed below), for which the CPU time for the calculations is proportional to N^2, where N is the number of atoms.

The distance constraints are used to set up a matrix of upper and lower distance bounds for the inter-atomic sub-structure distances. Undefined upper bounds are assigned a distance equal to the extended length of the protein. Undefined lower bounds are set equal to the sum of the relevant van der Waals radii.

Then, a procedure termed 'bounds smoothing' is undertaken, where the sub-structure's upper and lower bounds are adjusted to be consistent with respect to other inter-atomic distance. This is perfomed using the 'triangle inequality'. Consider three atoms a, b and c:

i.e.



The maximum distance between atoms *a* and *c* occurs in the colinear arrangement, *abc*, and cannot be greater than the sum of upper bound *ab* and upper bound *bc*

i.e. $ac_{upper} \leq ab_{upper} + bc_{upper}$

The minimum distance between atoms *a* and *c* occurs in a different colinear arrangement, *bca*, and cannot be smaller than the difference between lower bound *ab* and lower bound *bc*

i.e. $ac_{lower} \geq ab_{lower} - bc_{lower}$

For atoms which were not stereo-specifically distinguishable on NMR spectral analysis (*e.g.* as could be the case with Leucine HB1 and HB2), a pseudo-atom is placed at the mean position in order to represent them, and the distance constraints are adjusted accordingly.

After the bounds smoothing stage, a set of trial distances lying between the upper and lower bounds are randomly generated from a distribution that is biased towards the lower bounds (an unbiased distribution is weighted unrealistically towards larger distances, and therefore leads to extended structures). An ensemble of different starting distances gives rise to a set of final structures.

The next stage, termed as 'embedding', involves the conversion of the trial distances to coordinates. This involves the use of the metric matrix method to derive the metric matrix, M (with values m_{ij}), from the pairwise distance matrix, D (with values d_{ij} between atoms i and j). The values of M are given by:

$$m_{ij} \quad = \quad -\tfrac{1}{2}\left(d_{ij}^2 - d_{i\bullet}^2 - d_{j\bullet}^2 - d_{\bullet\bullet}^2\right)$$

where i corresponds to the row number, j corresponds to the column number, $d_{i\bullet}$ is the mean of each value in row i of D, $d_{j\bullet}$ is the mean of each value in row j of D, and $d_{\bullet\bullet}$ is the mean of all the elements in D. Matrix M is solved to find eigenvectors, w_k, and their corresponding eigenvalues, λ_k. The three largest eigenvalues correspond to the three dimensions of Cartesian space, denoted by k. The coordinate set is defined by vector points, v_k, which is given by:

$$v_k \quad = \quad \lambda_k^{\frac{1}{2}} w_k$$

Finally, it is necessary to perform a chirality check, since both the generated sub-structure and its mirror image will satisfy the distance constraints set. Assuming the sub-structures are well-defined, the correct chiral form may be determined as that which possesses the lower sum of improper angle energies. If a reference structure is available, the correct chiral form of the generated sub-structure may be defined as that with lower rms difference to the reference structure.

The optmisation procedure is performed after the sub-structures have been expanded into all-atom structures; the newly added atoms are initially arranged so that the regions of protein that they occupy are in extended conformation.

2.2.3 Optimisation of Structures: *XPLOR*

The loss of information from using only the three largest eigenvalues at the embedding stage of metric matrix distance geometry, and the lack of information due to an invariably non-exhaustive NMR constraints set, require that the generated 3D structures be optimised. This is perfomed using an initial energy minimisation stage, followed by molecular dynamics performed as a simulated annealing process, and finally by another energy minimisation stage.

The protein is described in terms of a forcefield, generally comprising of bond length, bond angle, dihedral angle, improper angle, van der Waals and NOE constraint terms.

i.e. $E_{tot} = E_{bond} + E_{angle} + E_{dihed} + E_{improp} + E_{vdW} + E_{NOE}$

The van der Waals term uses a simplified quartic repulsive term. For the application presented in this thesis, a square-well potential is used for the NOE term; that is, there is no energy contribution if the relevant atomic separation, d_{ij}, is within the distance bounds; however, a quadratic potential is applied for any distance outside these bounds using an appropriate force constant, k_{NOE}:

$$
\begin{aligned}
E_{NOE} \quad = \quad & k_{NOE}\left(d_{ij} - d_{upper}\right)^2 && \text{if } d_{ij} > d_{upper} \\
& k_{NOE}\left(d_{lower} - d_{ij}\right)^2 && \text{if } d_{ij} < d_{lower} \\
& 0 && \text{if } d_{lower} \leq d_{ij} \leq d_{upper}
\end{aligned}
$$

The dihedral angle term is subdivided into separate terms to include not only a general dihedral potential, but also ϕ and ψ constraints described by square-well potentials for secondary structural regions (*i.e.* α-helix and β-strand). Electrostatic (commonly omitted in such simulated annealing studies) and vicinal coupling constant (which could not be measured experimentally) terms were not included for the application in this thesis. Hydrogen bonds are incorporated as distance restraints using the NOE term.

Additionally, Cα,β secondary chemical shift (*i.e.* $E_{shift,C\alpha,\beta}$) and Hα proton chemical shift (*i.e.* $E_{shift,H\alpha}$) terms (Kuszewski *et al.*, 1995a and 1995b, respectively) are included within the forcefield as a part of the final stage of the refinement process for the application presented in this thesis. The energy terms are a function of the difference between experimental and expected shifts. The expected shifts are derived from empirical data from proteins for which the relevant chemical shifts were known and for which high resolution crystal structures were available. The Cα,β secondary and Hα proton chemical shift terms were expressed as harmonic potentials.

Ramachandran restraints (Kuszewski *et al.*, 1996) are also included for the refinement process, and serve to restrain dihedral angle values that are most likely to occur within proteins, thus improving the overall packing and stereochemical quality of generated structures. The energy term is proportional to the logarithm of probabilities of dihedral angles based on observed distributions (the data set was that of PROCHECK, Laskowski *et al.*, 1993; see section 2.3.1). These dihedral angles may be represented as one-dimensional (*i.e.* χ_1, χ_2, χ_3, χ_4) or two-dimensional (*i.e.* ϕ/ψ, χ_1/χ_2) grids of 8° per grid from which the information is taken.

For the initial energy minimisation, the van der Waals and NOE force constants are assigned low values; in the case of the van der Waals term this is to avoid large repulsive forces between atoms that are too close together, and in the case of the NOE term this is to accommodate atoms which are initially much too far apart. The main objective of initial minimisation is to ensure correct covalent geometry. The molecular dynamics is performed as a simulated annealing process over a picosecond timescale in steps of approximately one femtosecond. The temperature of the system is initially set high (*e.g.* 1000K) and the van der Waals and NOE force constants are gradually increased over consecutive molecular dynamics steps until their final values are reached at this temperature. Subsequently, the system is cooled slightly (*e.g.* by 25K) after each molecular dynamics step, finally reaching room temperature (*i.e.* 300K). Such a simulated annealing approach allows the system to overcome local maxima and cluster towards a global energy minimum. The objective of the final energy minimisation is to commit the structure to the minimal energy position of its current potential well.

The optimised structures are tested for consistent NOE constraint violations (*i.e.* usually ≥ 1.0Å distance violation in the first instance, and this is subsequently reduced to a distance violation of ≥ 0.5Å) across the whole ensemble. Any consistently violated NOE restraints are examined with reference to the NMR spectra, and the constraints are updated as necessary. The structure generation procedure is performed again beginning with metric matrix distance geometry. The calculation procedure is repeated iteratively until consistent NOE violations are no longer present.

2.3 ASSESSMENT OF QUALITY OF GENERATED STRUCTURES

After a 3D protein structure has been generated, the quality of the structure should be assessed, since both a structure derived from experimental data and a model predicted from mapping sequence to structure may contain structural imperfections and errors. Their identification not only highlights poorly defined structural regions, for example, but can also be used as a focal point or reference for structural improvement.

2.3.1 *PROCHECK* and *PROCHECK-NMR*

The program PROCHECK (Laskowski *et al.*, 1993) is used to assess the quality of a 3D protein structure. The criteria used are based on stereochemical considerations alone, and therefore the quality assessment is performed directly on the Cartesian coordinate structures. The associated program, PROCHECK-NMR (Laskowski *et al.*, 1996), may be used to test the quality of an ensemble of 3D models or NMR structures.

The data set of 'ideal' values was derived from 163 non-homologous (≤ 35% sequence identity) 3D protein structures solved by X-ray crystallography to a resolution[22] of 2.0Å or better and an R-factor[23] of 20% or less.

[22] Crystal planes and axes exist in reciprocal space. A crystal structure with a resolution of 2Å corresponds to data measurable up to ½Å^{-1} from the origin on a diffraction map. The resulting structural accuracy is higher (since more information is available) than for a crystal structure with 3Å resolution.

[23] The initial 3D protein structure must undergo refinement (*e.g.* energy minimisation, molecular dynamics) until the best agreement is found between the observed structure factor values, F_{obs}, and the values back-calculated from the electron density of the deduced structure, F_{calc} (the structure factor is a mathematical description of a diffracted X-ray from a particular crystal plane). The R-factor is a measure of disagreement between the two structures, and is given by $R = \Sigma|F_{obs}-F_{calc}| / \Sigma|F_{obs}|$ summed over all planes.

Some examples of the assessments employed within these programs are bond distances, bond angles, planarity (*i.e.* peptide bonds, aromatic rings, and sidechain endgroups for Arg, Asp, Asn, Glu, Gln), Ramachandran plots (identifying residues in the regions of the plot that are favoured to those that are disallowed; the number of residues in the most favoured regions should ideally be ≥90%), sidechain dihedral plots of χ_1 against χ_2 (identifying where individual residues lie with respect to the 'ideal' positions as found empirically), per-residue Ramachandran plot favourability, and per-residue secondary structure assignment (extracted from the 3D coordinates using the method of Kabsch & Sander, 1983).

Also specified are measures of overall quality for the query structure shown against corresponding empirically-determined values with respect to resolutions of the data set proteins. Examples of these are percentage of residues in most favoured regions of the Ramachandran plot, standard deviation of peptide bond planarity, bad contacts (≤2.6Å interatomic distance for non-hydrogen atoms) per 100 residues, and standard deviations of sidechain dihedral angles from 'ideal' conformational positions (e.g. χ_1 gauche plus, χ_1 gauche minus, χ_1 trans).

Since PROCHECK-NMR focuses on ensembles of structures, it additionally calculates measures of quality such as the circular variance of dihedral angles (*i.e.* ϕ, ψ, χ_1, χ_2, ϕ/ψ and χ_1/χ_2) across the whole ensemble. The circular variance, v, is defined as:

$$v \;=\; 1-\left[\left(\sum \cos\theta_i\right)^2 - \left(\sum \sin\theta_i\right)^2\right]^{\frac{1}{2}} / n$$

where θ_i is the dihedral angle and n is the number of members in the ensemble. The value of the circular variance ranges from 0 to 1, where a tighter clustering of dihedral angle values about its mean tends towards a circular variance of zero.

The G-factor is another measure of quality in both PROCHECK and PROCHECK-NMR for particular stereochemical parameters. It is a logarithm of probabilities based on the empirically observed distributions of the property in question. A higher G-factor value for the ϕ/ψ Ramachandran plot, for example, corresponds to residues being in higher-probability, or more ‘allowed’, mainchain conformational space. In the case of PROCHECK-NMR, the G-factor for a particular property is reported for that across the whole ensemble. Ideally G-factor scores should be above -0.5; values below -1.0 may need investigation.

2.3.2 3D Profiles

The 3D profile method of protein 3D structure assessment (Bowie *et al.*, 1991; Luthy *et al.*, 1992) is based on the representation of a 3D protein structure in terms of the structural environment of each amino acid in the sequence of that protein.

The environment for a particular amino acid is defined based on (1) the buried surface area of the residue, (2) the fraction of the sidechain surface area covered by polar atoms, and (3) the local secondary structure. These three parameters are used to allocate a particular residue in the sequence to one of 18 environment classes. Each residue is given a (logarithm of probabilities) score that reflects the compatability of that amino acid for that environment, based on observed frequencies within known 3D protein structures.

An overall score is then calculated for the protein model. A misfolded protein model yields a lower score since it possesses residues in environments with which they are not statistically compatible, and such models cannot often be detected as misfolded by potential energy (force field) methods.

A 3D profile can also be used to detect regions of incorrectly assigned residues by plotting the score against sequence number; any residues for which the score falls significantly low relative to the average score should be investigated to check whether the model is in error in that particular region.

The Verify3D[24] structure evaluation server performs such a calculation on-line for a 3D protein structure in PDB format.

2.4 MOLECULAR INTERACTION

This section describes some modelling techniques that allow the study of how proteins interact.

2.4.1 Electrostatic Surface Potentials

The electrostatic fields in and around proteins play a significant role in their function. A protein in solution may be considered to consist of two distinct phases; that is, the protein phase and the solvent phase, each with its own dielectric constant. Thus, a Coulombic treatment of the protein system is not appropriate.

[24] URL http://www.doe-mbi.ucla.edu/verify3d.html

The Poisson-Boltzmann equation, which takes into account a spatially varying dielectric function, is given by

$$\nabla \cdot [\varepsilon(x) \nabla \phi(x)] - \kappa(x)^2 \sinh(\phi(x)) = -4\pi \rho(x)$$

where $\phi(x)$ is the electrostatic potential to be calculated, $\varepsilon(x)$ is the spatial dielectric function, $\kappa(x)$ is the Debye-Huckel parameter (taking into account the electrostatic effects of salt which is present in real biological systems), and $\rho(x)$ is the charge distribution function.

The hyperbolic sine term, *sinh(*ϕ*(x))*, in the Poisson-Boltzmann equation describes the nature of salt accumulation over regions of the protein with respect to electrostatic potential (assuming the salt is a 1:1 charge pair). The first term in the mathematical series expansion of *sinh(*ϕ*(x))* gives $\phi(x)$; thus, $\phi(x)$ is approximately equal to *sinh(*ϕ*(x))* when $\phi(x)$, the electrostatic potential, is small (and, therefore, when the salt concentration is low). This gives rise to the linearised Poisson-Boltzmann equation, given by:

$$\nabla \cdot [\varepsilon(x) \nabla \phi(x)] - \kappa(x)^2 \phi(x) = -4\pi \rho(x)$$

Its analytical solution is feasible for only a few idealised shapes (*e.g.* spheres, cylinders). The complex surfaces formed by biological molecules, such as proteins, require that the linearised Poisson-Boltzmann equation be solved numerically.

2.4.1.1 DELPHI

DELPHI (Gilson *et al.*, 1987; see also Nicholls & Honig, 1991) is a program that calculates the electrostatic potentials of molecules using a numerical method to solve the linearised

Poisson-Boltzmann equation. The protein is placed into a cubic lattice of 65 x 65 x 65 grid points, and the electrostatic potentials are assigned to these gridpoints.

The reduction of the linearised Poisson-Boltzmann equation to finite difference form is given by:

$$\phi_o = \frac{\left(\sum_{i=1}^{6} \varepsilon_i \, \phi_i\right) + 4\pi q_o / h}{\left(\sum_{i=1}^{6} \varepsilon_i\right) + (\kappa_o \, h)^2}$$

where ϕ_o is the potential at a particular gridpoint, ϕ_i is the potential at the six nearest gridpoint neighbours, *h* is the grid spacing (Å), q_o is the charge assigned to the gridpoint (mapped as a distance-determined proportion of nearby atom charges within a given cutoff distance), ε_i is the dielectric constant assigned at the grid midpoints, and κ_o is the Debye-Huckel parameter for a particular gridpoint (zero if not in salt).

The dielectric constant, ε_i, is set to one value for the protein interior (*i.e.* $\varepsilon_{protein}$, default=2) and another value for bulk solvent (*i.e.* ε_{water}, default=80). The protein surface is defined by the solvent accessible surface, using a spherical probe with radius approximately equal to that of water (*i.e.* r_{water}, default=1.8Å). Thus, grid midpoints within the protein are assigned with dielectric constant equal to $\varepsilon_{protein}$.

The Debye-Huckel parameter, κ_o, is assigned to zero if the gridpoint is within the salt exclusion zone. This exclusion zone is defined as being inside the protein's van der Waals surface, and within a salt atom radius (default=2Å, *i.e.* sodium) outside the protein's van der Waals surface. The Debye-Huckel parameter is otherwise assigned as:

$$\kappa_o = \left(\frac{8\pi e^2 N_A I}{1000 k_B T} \right)^{\frac{1}{2}}$$

where e is the charge on an electron, N_A is the Avagadro's number, I is the bulk salt concentration[25], k_B is the Boltzmann constant, and T is the temperature.

The potential at the boundary of the grid can be assumed to be equal to zero if the grid boundary is far from the protein molecule. This corresponds to the protein being placed in a large grid. However, since a grid of a given size (*i.e.* 65 x 65 x 65 gridpoints) is used, then making the grid larger relative to the molecule leads to a coarser representation of the protein surface with respect to the electrostatic potentials. Increasing the number of gridpoints would place a greater demand on computer memory. Thus, a 'focusing' approach (McAllister *et al.*, 1985) may be used for increasing the accuracy of the electrostatic potential calculations. For this approach, the ratio between the largest linear dimension of the protein and the dimension of the side of the grid (expressed as a percentage) is progressively increased over successive runs (*e.g.* the user might specify 20%, 50%, then 80%). The boundary gridpoint potential values for each successive run are assigned as those of the output of the previous run (in the first run these are set to zero).

Even with the focusing approach, significant errors close to charges and the dielectric boundary may occur in some cases. These errors depend on the precise position and orientation of the protein system relative to the grid. Therefore, a 'rotational averaging' approach is implemented. That is, for each run, the calculation is repeated a number of times

[25] Strictly speaking, this should be the ionic strength of the solvent, given by $I=\frac{1}{2}\Sigma(c_iZ_i^2)$ where c_i is the salt concentration, Z_i is the charge assignment of an individual ion, and where the summation extends over each salt ion type in the solution. However, assuming that the salt comprises only NaCl (*i.e.* singly charged ions of a 1:1 electrolyte), then the ionic strength is $I=\frac{1}{2}[c(1)^2 + c(-1)^2]$ which simplifies to the concentration, c.

with the protein rotated by different angles with respect to the grid, and the average of the results taken.

2.4.1.2 GRASP

GRASP (Graphical Representation and Analysis of Structural Properties; Nicholls *et al.*, 1991) is a software package that is useful for displaying and analysing electrostatic potentials on protein surfaces.

The protein coordinates are read in from a PDB file, and may be represented in a number of ways. The protein surface may either be calculated as a molecular surface or a solvent accessible surface; the solvent accessible surface describes the protein surface in terms of a water probe (default radius of sphere = 1.4Å) mapped over the molecular surface. The surface comprises of triangles generated using a grid-based method, and is usually represented by solid rendering.

An electrostatic potential grid map may be read in from the output of a DELPHI run (*i.e.* the '*.phi*' file) or calculated by GRASP using a similar but simpler version to DELPHI (the former method is more precise, and therefore the method of choice). The electrostatic potentials from the grid are interpolated to the protein surface, and the display is contoured either to absolute values or to within a user-specified range.

The GRASP electrostatic potential calculation involves the use of two 33 cubed grids, one nested within the other. The size of the inner grid is set to be larger than the maximum dimension of the protein by the radius of one water molecule at either end. The outer grid is twice as big as the inner grid, and shares the same centre. The potentials of the outer grid are solved first, then interpolated and refined on the inner grid. Finally, all the potentials are mapped onto the gridpoints of a 65 x 65 x 65 grid which occupies the same volume as the outer grid.

Field lines may be calculated for the protein surface after the electrostatic potentials have been read in. The user specifies the number of seed points (*i.e.* starting points) on the surface or within a user-scribed surface region, and the field direction is calculated and displayed for a positive or negative charge from each point.

The dipole moment of the protein is calculated from the charges on the atoms[26,27]. The charge sum and average charge-weighted position is calculated for both positive and negative charges. The magnitude of the dipole moment is determined by the smaller charge sum multiplied by the distance between the charge-weighted centres. The direction of the dipole is from the negative to positive charge centre; the graphically displayed dipole moment vector starts at the average of the charge-weighted centres. The monopole is equal to the net charge of the atoms.

2.4.2 Ligand Binding

The binding of ligands may be predicted for proteins with either modelled or experimentally determined 3D protein structures. This is perfomed by identifying the binding site of the

[26] The GRASP charge library specifies charges over the minimal set of 'relevant' atoms only (*e.g.* charge=1.0 for Lys NZ). The charges read in from the DELPHI output PDB file (taken from the DELPHI charge library) exist as partial charges spread more 'comprehensively' over the molecule.
[27] For two equal and opposite charges, $+q$ and $-q$, and separated by distance, r, the magnitude of dipole moment, μ, is given by $\mu=qr$.

protein under consideration, then characterising the binding of the ligand within that site. For each of these steps, an appropriate program is presented below.

2.4.2.1 Identification of Protein Binding Site: SURFNET

SURFNET (Laskowski, 1995) is a program that can be used for generating van der Waals surfaces and determining 'gap regions' within proteins. The identification of cavities is useful for determining binding sites within proteins, and is achieved as follows.

The 3D structure of a query protein is first placed in a grid of user-defined size and grid separation. A sphere is then placed between each pair of atoms in the protein structure, the sphere centre being the midpoint between the two atoms. The sphere radius is defined as maximal when it comes into contact with the van der Waals radius of any atom, and its minimum is defined by a pre-determined cutoff (usually 1Å).

Gridpoints around each sphere centre are assigned scores that correspond to a Gaussian function of the form:

$$\rho = \rho_o e^{-kr^2}$$

where ρ is the density, or score, assigned to a given gridpoint at distance r from the sphere centre, ρ_o is the score at the sphere centre (set to 200), and k is defined by:

$$k = \frac{\ln 2}{r_{sphere}^2}$$

so that the score at the radius of the sphere (*i.e.* 100) is half the score at its centre. Only gridpoints within a distance of $3r_{sphere}$ radii are scored. To prevent the distortion of a particular gridpoint score due to contributions from more than one sphere, only the largest score contribution is taken at each gridpoint.

The gap regions are defined by a user-specified gridpoint cutoff score (usually 100). Note that to generate the van der Waals surface, the spheres are centred at the atom centres, and r_{vdW} is used for a given atom in place of r_{sphere} above; thus, the van der Waals surface of the protein is defined by a contour level of score=100.

The cavities (or molecular surface) can be viewed using molecular graphics software such as InsightII (Biosym, San Diego, USA). Particular cavities may be selected for display using the *mask* routine. SURFNET also provides approximate volumes based on the number of gridpoints within the cavities. The largest cavity corresponds to the active site of an enzyme in the vast majority of cases (Laskowski *et al.*, 1996b).

2.4.2.2 Interaction of Ligand with Protein: GRID

GRID (Goodford, 1985), calculates the interaction energy between a protein and specific probe types. Thus, the most likely binding mode between a protein and its ligand may be determined, where probes are used to represent the chemical nature of the ligand. The probes correspond to functional groups (*e.g.* methyl group, carboxylate oxygen).

The protein is placed in a user-defined grid and the energy values are assigned at the gridpoints. The interaction energy at each sampled position is given by the sum of the Lennard-Jones potential, the hydrogen bond potential and the electrostatic potential; that is,

$$E_{xyz} = \sum E_{LJ} + \sum E_{HB} + \sum E_{el}$$

The Lennard-Jones term is given by:

$$E_{LJ} = \frac{A}{d^{12}} - \frac{B}{d^{6}}$$

where *d* is the pairwise distance between non-bonded atoms, and *A* and *B* are constants.

The hydrogen-bond potential is a 6,4-term rather than the more frequently-used 12,10-term producing a broader energy minimum, therefore resulting in a less volume-limited spatial interaction. The hydrogen-bond is nevertheless treated as being orientation dependent. The H-bond term is given by:

$$E_{HB} = \left[\frac{C}{d^{6}} - \frac{D}{d^{4}}\right]\cos^{m}\theta$$

where *d* is the pairwise distance between non-bonded atoms, *C* and *D* are constants, *m* is usually equal to 4, and θ is the angle *MHP*, where *M* is the protein H-bond donor atom, *H* is the protein hydrogen atom and *P* is the probe accepting the H-bond. If $\theta \leq 90^{\circ}$, the $cos^{m}\theta$ term is set to zero, since the orientation is not conducive to a H-bond. If the H-bond donor is the probe group, the $cos^{m}\theta$ term is set to unity, since it is assumed that the probe can orient itself in order to form the most effective H-bond interaction with the protein's acceptor atom.

For the electrostatic potential, the system is treated as existing with two distinct phases separated by a planar interface: a protein phase with dielectric constant ζ=4, and a solvent phase (water) with dielectric constant ε=80. The depth of any point within the protein phase is defined by the number of protein atoms within a vicinity of 4Å. The electrostatic term is given by:

$$E_{el} = \frac{pq}{K\zeta}\left[\frac{1}{d} + \frac{(\zeta - \varepsilon)}{(\zeta + \varepsilon)(d^2 + 4s_p s_q)^{\frac{1}{2}}}\right]$$

where p is the charge on probe, q is the charge on protein, K is a constant, d is the probe to atom distance, ζ is the protein dielectric constant, ε is the solvent dielectric constant, s_p is the depth of probe within protein phase, s_q is the depth of protein atom within protein phase. The term $4s_ps_q$ is set to zero if the probe has less than 7 protein atoms within a distance of 4Å, since this gives the appropriate functional form of the electrostatic potential when the probe is outside the protein phase.

The favourable regions for each probe can be displayed as contour surfaces with respect to the protein using molecular graphics software (*e.g.* InsightII). This gives an indication of the favourable binding positions between the different probes (and therefore the ligand under consideration) and the protein.

2.4.3 Characterisation of Channels

The characterisation of channels within proteins and the determination of their minimum radial dimensions may aid in illustrating how some proteins accommodate substrate access into their (buried) binding sites, and also offer structural insights into the internal surfaces of ion channel proteins.

2.4.3.1 HOLE

HOLE (Smart *et al.*, 1993) is a program that conceptually traces the path of a sphere with maximal but flexible radius through a particular channel in a protein. The channel is represented as an irregular tube defined by the locus of the outer surface of the sphere as it squeezes through the protein.

The procedure begins with the user-specification of the coordinates of an initial point, *p*, that lies anywhere within the protein channel, and a vector, *v*, which is approximately in the direction of the channel.

The point *p* is considered to be on a plane that is perpendicular to the direction of vector *v*. A sphere, with its centre on this plane, is assigned maximum radius at the point of contact with the van der Waals radius of any protein atom over 3D space (any ions or solvent molecules are excluded from consideration). A Monte Carlo simulated annealing procedure is performed over 1000 steps to find the largest sphere for which the centre lies on the plane.

Next, a small displacement is made in the direction of the approximate user-specified direction vector, *v*, in order to calculate the optimal position of a sphere centre on the new plane. The process is continued on successive planes until the end of the channel has been reached, that is, until the sphere radius exceeds 5Å. Then, the whole process is repeated in the direction -*v* in order to ensure that the full channel is characterised.

The channel may be visualised on molecular graphics packages including InsightII as a plot file in conjunction with the protein under study.

2.4.4 Electron Transfer

In weakly coupled donor-acceptor systems, the rate of electron transfer, k_{et}, is given by:

$$k_{et} = \frac{4\pi^2}{h}|V_R|^2(F.C.)$$

where V_R is the electronic coupling between donor and acceptor moieties and contains information on the nature of the medium and the donor-acceptor separation, *R*; *(F.C.)* is the Franck-Condon factor which is determined by the nuclear processes associated with the electron tunnelling.

Marcus Theory (Marcus, 1956), formulated by the classical treatment of the previous equation, states that the rate of electron transfer is dependent on the free energy of the reaction $-\Delta G^o$, the reorganisation energy of the system, λ, as well as the donor-acceptor separation and transfer medium. The Marcus expression is given by:

$$k_{et} = \left(\frac{\pi}{h^2\lambda k_B T}\right)^{\frac{1}{2}}|V_R|^2\exp\left[\frac{-(\Delta G^o-\lambda)^2}{4\lambda k_B T}\right]$$

where k_B is the Boltzmann constant and T is the temperature.

Since an orbital wavefunction decays exponentially with respect to distance, it is expected that k_{et} also decreases exponentially with respect to the distance, *R*, between donor and acceptor moieties:

$$k_{et} = k_o e^{-B(R-R_o)}$$

where k_o is the value for k_{et} at donor-acceptor contact distance, R_o, and B is the rate of exponential decay which also describes the effect of the medium on the reaction.

2.4.4.1 PATHWAYS

The program PATHWAYS (Beratan *et al.*, 1991) calculates the most likely path of electron transfer from user-defined starting point and endpoint bonds. The electron tunnelling is represented as discrete spatial steps from one bond orbital to another (including that between a lone pair of electrons and its associated atom), and consists of both through-bond and through-space progression.

The coupling, ε, between each interacting orbital that constitutes the complete electron transfer pathway is expressed as:

$$\varepsilon \;=\; \varepsilon_o e^{-\beta(r-r_o)}$$

where ε_o is the value for ε at orbital contact distance, r_o, and β is the rate of coupling decay with respect to distance between bond orbitals, *r*. The coupling decay factor, β, was defined for through-bond (including hydrogen bonds) and through-space coupling from empirical studies (Beratan *et al.*, 1991).

The PATHWAYS program first assigns hydrogen atoms and electron lone pairs to the protein (which may be viewed in an output file in PDB format). Each atom in the protein is then assigned to a particular configuration class (*e.g.* sp2LL; that is, an atom with sp^2 hybridisation and with two lone pairs of electrons), and also to an atom class (*e.g.* ring). Atom classes are then classified into atomic orbital classes (*e.g.* sigma). Interactions between atomic orbital classes are then assigned into orbital coupling classes which can be either through-space or atom-mediated (*e.g.* H-bond).

The tunnelling pathway is defined as the combination of orbital couplings linking the donor and receptor moieties. The best path is that with the maximum coupling pathway, that is, the path with highest coupling values in consideration of both through-bond and through-space jumps.

CHAPTER 3

MODELLING THE 3D STRUCTURE OF NADPH - CYTOCHROME P_{450} REDUCTASE (P450R)

3.1 Introduction

Human NADPH-cytochrome P_{450} reductase (P450R) is an enzyme that exists in the cytosol of most eukaryotic cells. It binds to the cell's endoplasmic reticulum in order to catalyse the transfer of an electron from NADPH (Nicotinamide Adenine Dinucleotide Phosphate in reduced state), a cellular electron carrier, to cytochrome P_{450}, a heme-containing protein which is also bound to the endoplasmic reticulum. The cytochrome P_{450} is responsible for the oxidation of endogenous and foreign substances, including drugs and steroids.

The P450R sequence, consisting of 677 residues, may be represented by the relative location of the domains (see Figure 3.1 in methods section of this chapter) as presented by Porter (1991).

The P450R enzyme makes use of two prosthetic groups to facilitate its functionality, namely FMN (Flavin MonoNucleotide) and FAD (Flavin Adenine Dinucleotide). The path of the electron begins from the nicotinamide group of NADPH, it then proceeds to the isoalloxazine ring (also known as the flavin ring) in the enzyme's FAD group, then to the isoalloxazine ring in the enzyme's FMN group, and finally to the heme group of cytochrome P_{450} (Vermilion *et al.*, 1981).

In this study, the 3D structure of the P450R was modelled in order to rationalise the functionality of this enzyme. This was undertaken in collaboration with Prof. G.C.K. Roberts and co-workers, Department of Biochemistry, University of Leicester.

A crystal structure for P450R became available (after the modelling presented in this chapter was performed) excluding the membrane binding domain (PDB code 1AMO; Wang *et al.*, 1997). The domains are divided similar to that described in Figure 3.1. The coordinates of the crystal structure are on hold until 17 June 1998, and therefore no direct 3D structural comparison can be made with the predicted structures. However, a secondary structural comparison may be made and is illustrated in the sequence alignment figures (*i.e.* Figures 3.3, 3.4, 3.5 and 3.9).

Also available subsequent to the modelling presented here were the crystal structure (Zhao *et al.*, 1996) and the NMR structure (Barsukov *et al.*, 1997) of the FMN domain of P450R. The NMR structure was based on the homology model presented in this chapter.

3.2 Methods

Figure 3.3 describes the method of 3D template identification for each domain. That is, whether it was modelled based on sequence similarity (*i.e.* homology; therefore, more likely to be correct) or by the propensity to take up a particular fold (*i.e.* threading; therefore, less likely to be correct). Homology modelling is the method of choice based on accuracy of model.

	N-terminal					C-terminal
Domain	Membrane binding domain	FMN binding domain	Linker region	FAD binding domain	'Insertion' domain	FAD/ NADPH binding domain
No. of Residues	76	152	38	63	117	231
Method of Template Identification	Threading	Homology	N/A	Homology	Threading	Homology

Figure 3.1 Representation of P450R in terms of structural domains, including method of identification of template structures for each domain for modelling purposes.

The flavin-containing domains (*i.e.* the domains directly associated with electron transfer) were modelled based on homology; it was then attempted to dock these domains together in order to gain an insight into the structure-function relationship of the enzyme. The linker region was uncharacterised in terms of 3D structure from homology- and threading-based studies.

The methodological approach to this study is described by the flowchart in Figure 3.2. The methods are detailed in chapter 2.

Sequence similarity searches were originally performed using the BLAST algorithm (Altschul *et al.*, 1990), the output of which was used as a basis for 3D modelling; another sequence similarity search program, FASTA (Pearson & Lipman, 1988; also Pearson, 1990), produced very similar results to BLAST.

Threading calculations were required for portions of the P450R sequence which did not exhibit sequence homology with known 3D protein structures (see results and discussion

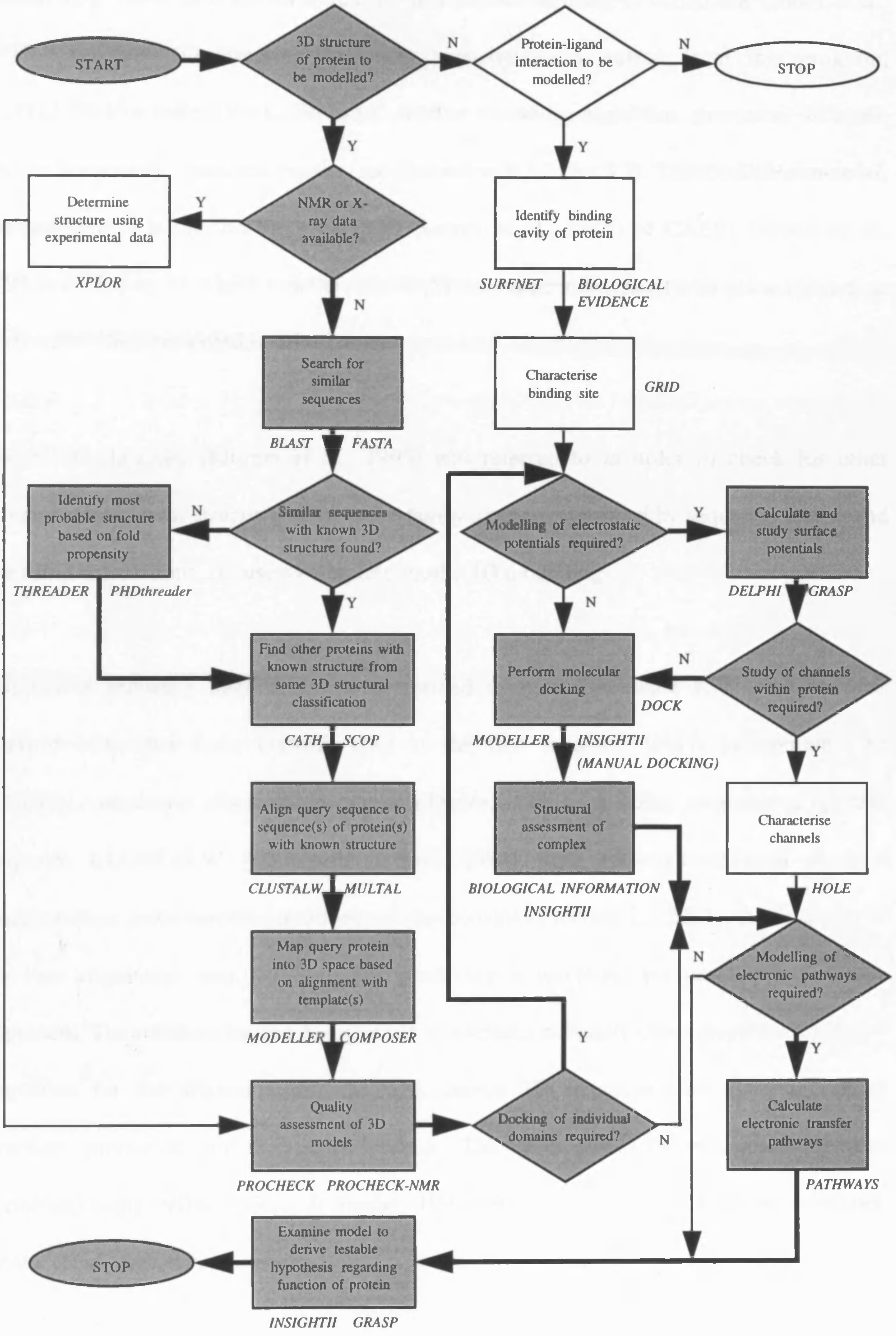

START
3D structure of protein to be modelled?
N
Protein-ligand interaction to be modelled?
N
STOP
Y
Y
Determine structure using experimental data
Y
NMR or X-ray data available?
Identify binding cavity of protein
XPLOR
SURFNET
BIOLOGICAL EVIDENCE
N
Search for similar sequences
Characterise binding site
GRID
BLAST
FASTA
Identify most probable structure based on fold propensity
N
Similar sequences with known 3D structure found?
Modelling of electrostatic potentials required?
Y
Calculate and study surface potentials
THREADER
PHDthreader
Y
N
DELPHI
GRASP
Find other proteins with known structure from same 3D structural classification
Perform molecular docking
N
Study of channels within protein required?
DOCK
CATH
SCOP
MODELLER
INSIGHTII
(MANUAL DOCKING)
Y
Align query sequence to sequence(s) of protein(s) with known structure
Structural assessment of complex
Characterise channels
CLUSTALW
MULTAL
BIOLOGICAL INFORMATION
INSIGHTII
HOLE
Map query protein into 3D space based on alignment with template(s)
N
Modelling of electronic pathways required?
MODELLER
COMPOSER
Y
Y
Quality assessment of 3D models
Docking of individual domains required?
N
Calculate electronic transfer pathways
PROCHECK
PROCHECK-NMR
PATHWAYS
STOP
Examine model to derive testable hypothesis regarding function of protein
INSIGHTII
GRASP

Figure 3.2 Flowchart of Methods for Modelling P450R (shaded boxes)

section 3.3). These calculations were originally performed using THREADER (Jones *et al.*, 1992b) and the 3D structure modelling was based on output from this program; TOPITS/PHDthreader (Rost, 1995a,b), another threading algorithm, produced different, but not necessarily improved, results (see discussion in section 3.3). THREADER, however, performed well in the threading category (Lemer *et al.*, 1995) of CASP1 (Moult *et al.*, 1995; see Chapter 4), which was the only CASP assessment that had taken place at the time of the modelling presented in this chapter.

The SCOP database (Murzin *et al.*, 1995) was referred to in order to check for other proteins with known structure in the same family as those identified by sequence search and threading calculations, for use as templates in the 3D modelling.

Automated sequence alignment was performed using the software package Cameleon (Oxford Molecular Ltd., Oxford, UK) in the first instance, which incorporates the MULTAL sequence alignment program (Taylor, 1988). Another sequence alignment program, CLUSTALW (Thompson *et al.*, 1994), was subsequently used since it incorporates a more involved treatment of gap penalties (Section 2.1.2.2.1). A consensus of the two alignments was the basis for producing a resultant, yet unrefined, sequence alignment. The resultant sequence alignment was refined manually using graphical display of structures for the alignment proteins with known 3D structure, and using secondary structure prediction for the query protein. The secondary structure prediction was performed using PHDsec (Rost & Sander, 1993a; Rost & Sander, 1993b; Rost & Sander, 1994a; Rost *et al.*, 1994).

An ensemble of 3D models for the P450R structure was generated using MODELLER (Sali & Blundell, 1993), and was based on the proteins identified as structural templates using sequence similarity and threading approaches. A fragment-based approach, COMPOSER (Sutcliffe *et al.*, 1987a,b), was used initially for model generation. However, MODELLER possesses an expandable library of prosthetic groups, any of which can automatically be built into a protein during the model generation if suitable distance constraints are specified. Furthermore, COMPOSER generates a single model whereas MODELLER can generate an ensemble of structures, and cluster analysis based on pairwise RMS differences between a COMPOSER model and a MODELLER ensemble of models for a 'test' domain in P450R showed that the COMPOSER model seemed to exist as a structural 'outlier' with respect to the members of the MODELLER ensemble; therefore, it is not 'meaningful' to use the output structures of both the programs together for the purpose of selecting a representative structure. MODELLER was thus selected as the method of choice.

The selection of a representative structure from the MODELLER-derived ensemble of structures was perfromed using NMRCLUST (Kelley *et al.*, 1996). Quality assessment of structures was performed using PROCHECK (Laskowski *et al.*, 1993).

The modelling of P450R necessitated the subdivision of the protein into distinct domains (see Results and Discussion section 3.3). The docking of these domains for the purpose of defining interdomain interaction was attempted using both manual and automated methods.

The manual docking involved using both spatial and electrostatic surface potential complementarity. The electrostatic potentials were calculated using DELPHI (Gilson *et al.*, 1987; see also Nicholls & Honig, 1991) using a focusing approach where the protein occupied 20%, 50% and 80% of the grid volume in successive runs. The manual docking also involved spatial alignment of dipole moment vectors, which may signify the general direction of electron movement, and were calculated using GRASP (Nicholls *et al.*, 1991).

The automated docking method used was DOCK (Shoichet & Kuntz, 1991; not described in chapter 2) which treats the two separate molecules to be docked as either 'receptor', represented by spheres in cavities or grooves, and 'ligand', represented by spheres inside protruding areas. The size of 'receptor' and 'ligand' molecules is an important factor in the docking procedure. The spheres representing a large molecule are required to be clustered into smaller subsets of 25-60 spheres by suitable parameterisation of sphere size limits. Any of the sphere subsets, one from the 'receptor' and one from the 'ligand' may be selected as the regions to be docked. The docking is based on superposition of 'receptor' and 'ligand' spheres. DOCK allows geometry-based docking using sampling of space and restricted by steric overlap, and also allows docking based on geometry with the AMBER forcefield using energy minimisation. The program is highly parameter sensitive.

Both manual and automated docking methods treat the structures to be docked as rigid bodies, that is, it is assumed that binding does not distort the molecules significantly from the unbound conformation. However, the FMN and FAD+FAD/NADPH domains of P450R were docked using MODELLER (Sali & Blundell, 1993), a simulated annealing model generation approach based on user-specified inter-domain hydrogen bond restraints, for the purpose of electron pathway calculations which were performed using PATHWAYS (Beratan *et al.*, 1991).

3.3 Results and Discussion

P450R exhibits homology to continuous segments of the following proteins, all of which contain both FMN and FAD prosthetic groups: *Escherichia coli* Sulphite Reductase (SulR), Human Nitric Oxide Synthetase (NOS3) and the reductase domain of *Bacillus megaterium* Cytochrome P_{450} Reductase (BM3). However, none of these proteins have known 3D structures. Sequence homology of P450R with known 3D structures was not exhibited over the entire length of the P450R sequence. Therefore, 3D models were generated for sub-regions of P450R using the most suitable approach for each particular sub-region (illustrated in Figure 3.1 in methods section of this chapter). Interaction of these separate domains would then determine the overall structure and packing of the query protein.

The modelling of domains within P450R and the subsequent inter-domain docking are discussed in the sections below.

3.3.1 Threading-Modelled Domains of P450R

The domains which were modelled based on a threading approach are the membrane-binding and 'insertion' domains (refer to Figure 3.1). This section also includes the threading-based study of the 38-residue 'linker region'.

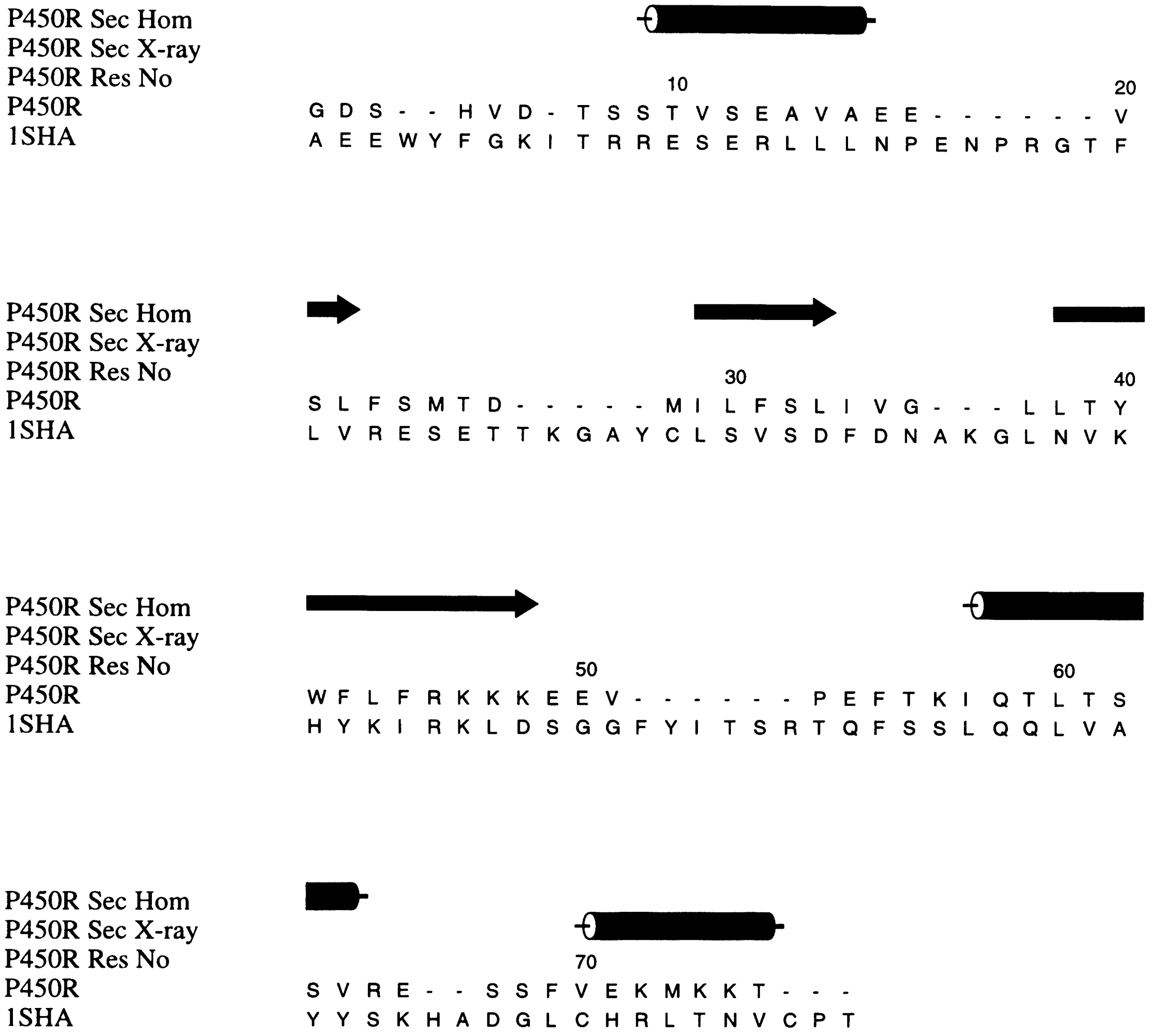
3.3.1.1 Membrane-Binding Domain

The 76 N-terminal residues is the endoplasmic reticulum membrane binding domain (Porter, 1991).

The THREADER (Jones *et al.*, 1992b) program assigned viral *phosphotransferase* (PDB accession code 1SHA) for the structural prediction of this domain. The significance score for the filtered pairwise energies is -3.35, which is categorised as 'significant - good chance of being correct' (DNA-binding regulatory protein, PDB code 1CMB, had the next highest score of -2.55, which is categorised as a 'poor score'). The significance score for the weighted sum of pairwise and solvation energy is -3.65 (a guideline score for significance in this case is less than -2.0). The predicted structure of this domain did not exhibit such α-helical characteristics that may be deemed to act as a membrane anchor. The *phosphotransferase* crystal structure is ligand bound to phospho-pentapeptide, and the 3D structure of this complex suggests that the membrane binding site of P450R should be located in the vicinity of the β-sheet face within the structure.

PHDthreader/TOPITS (Rost, 1995a,b) assigned rat prostatic acid phosphatase (PDB code 1RPA). The significance score, *ZALI*, is 1.72 which is categorised by TOPITS as having a 33% chance of being correct (purine nucleoside phosphorylase, PDB code 1ULA, had the next highest score of 1.70, which is also categorised as having a 33% chance of being correct).

Thus, a probable fold from one of these methods was not rated as a probable fold by the other method. The 3D structure of the membrane binding domain was modelled on 1SHA using the THREADER output alignment, which is shown in Figure 3.3. Comparison of the secondary structure of the identified fold and the secondary structure prediction of this region of P450R using PHDsec (Rost & Sander, 1993a; Rost & Sander, 1993b; Rost & Sander, 1994a; Rost *et al.*, 1994) did not give significant justification for refinement of any



```
P450R Sec Hom
P450R Sec X-ray
P450R Res No                             10                            20
P450R            G D S - - H V D - T S S T V S E A V A E E - - - - - - V
1SHA             A E E W Y F G K I T R R E S E R L L L N P E N P R G T F

P450R Sec Hom
P450R Sec X-ray
P450R Res No                                 30                        40
P450R            S L F S M T D - - - - - M I L F S L I V G - - - L L T Y
1SHA             L V R E S E T T K G A Y C L S V S D F D N A K G L N V K

P450R Sec Hom
P450R Sec X-ray
P450R Res No                       50                              60
P450R            W F L F R K K K E E V - - - - - - P E F T K I Q T L T S
1SHA             H Y K I R K L D S G G F Y I T S R T Q F S S L Q Q L V A

P450R Sec Hom
P450R Sec X-ray
P450R Res No                       70
P450R            S V R E - - S S F V E K M K K T - - -
1SHA             Y Y S K H A D G L C H R L T N V C P T
```

Figure 3.3 Threading-based alignment for the P450R membrane binding domain. Row 1 shows the secondary structure of the predicted model. Row 2 shows the secondary structure of the corresponding residues in the crystal structure of Wang *et al.* (1997). The crystal structure included only the last 12 residues of the above region of P450R sequence.

regions of the alignment. Also, there were no consistent structural violations in the 3D model produced by MODELLER.

The overall G-factor, as determined by PROCHECK (Laskowski *et al.*, 1993), for the representative model was 0.01 (ideally G-factor scores should be above -0.5; values below -1.0 may need investigation), and 78% of residues in this structure had their ϕ/ψ angles in the most favoured regions of Ramachandran space. This signifies satisfactory quality of stereochemical structure. However, the subsequently-solved crystal structure of P450R (Wang *et al.*, 1997) included the last 12 residues of this domain; the model does not contain a helix where it is present in the crystal structure (see Figure 3.3), thus suggesting an incorrect model.

3.3.1.2 Linker Region

A linker region of 38 residues exists between the FMN and FAD domains.

THREADER (Jones *et al.*, 1992b) identifies human interleukin 8 (PDB code 3IL8) as the most probable reference for the fold of this insertion domain. The significance score for the filtered pairwise energies is -2.93, which is categorised by THREADER as 'borderline significant - possibly correct' (steroid-binding protein, PDB code 1UTG, had the next highest score of -2.87, which is also categorised as 'borderline significant'). The significance score for the weighted sum of pairwise and solvation energy is -2.35 (a guideline score for significance in this case is less than -2.0).

TOPITS (Rost, 1995a,b) assigned human rhinovirus coat protein (PDB accession code 4RHV) as having the most probable fold of this domain. The significance score, *ZALI*, is 1.83 which is categorised by TOPITS as having a 33% chance of being correct (T-cell surface glycoprotein CD4 domains 3 and 4, PDB code 1CID, had the next highest score of 1.39, which is also categorised as having a 33% chance of being correct).

A probable fold from one of these methods was not rated as a probable fold by the other method.

The docking of the FMN- and FAD-containing domains (section 3.3.3) which aimed, in particular, to rationalise electron transfer in terms of the relative positioning of these domains, showed that the C-terminus of the FMN domain and the N-terminus of the FAD domain is spatially distant. If the model complex is correct, the linker region would have to bridge a gap of 25Å (Cα-Cα). The modelling of the linker region based on either of the threading results would not result in a suitable proximity between the prosthetic groups.

The linker region was thus uncharacterised, and not modelled. The subsequently-derived crystal structure (Wang *et al.*, 1997) indicates that this region exists as 'coil' except for a strand (residues 246-251) which is complementary to the single β-strand in the 'insertion domain' crystal structure (Wang *et al.*, 1997).

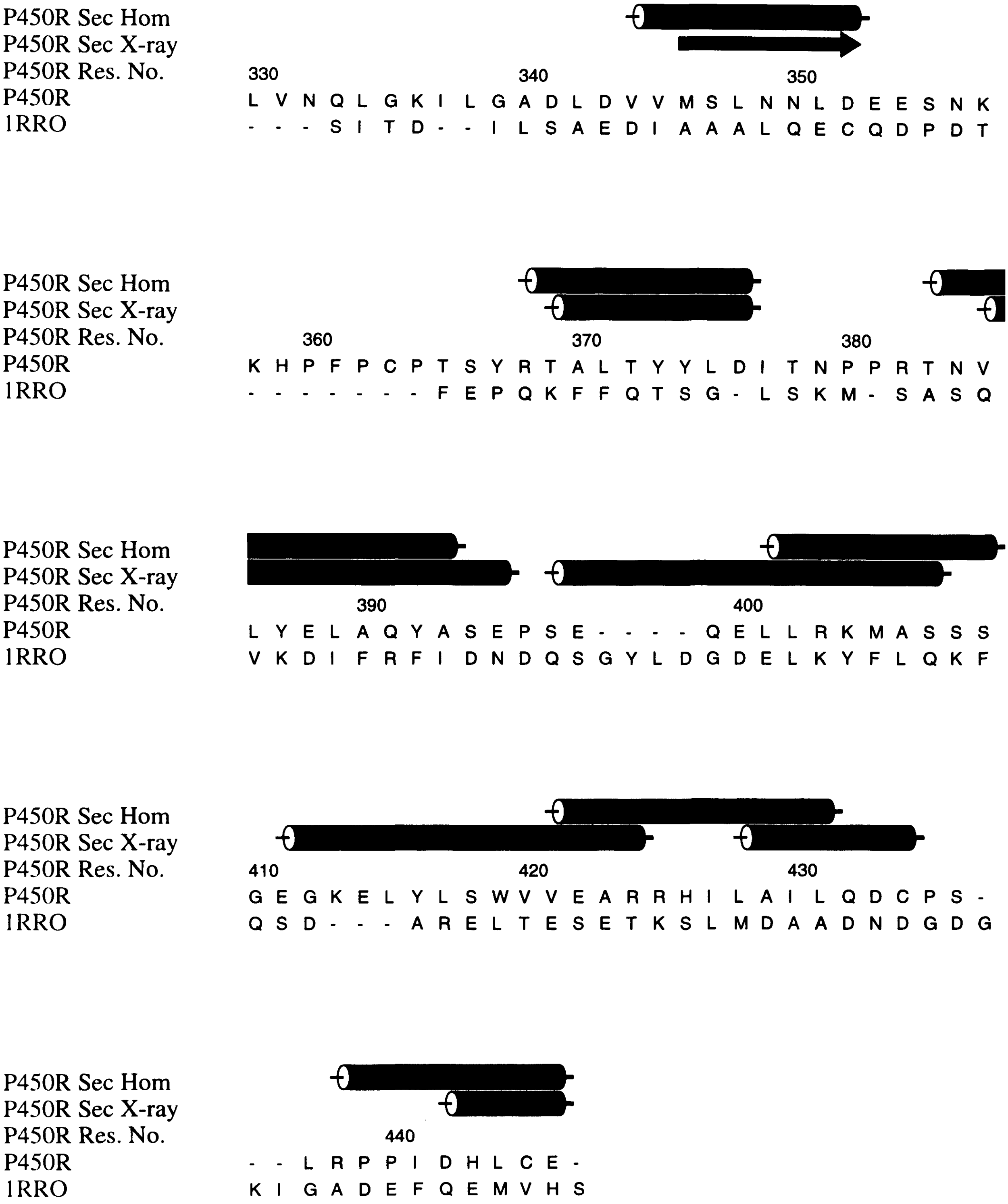
3.3.1.3 'Insertion' Domain

An 'insertion' sequence of 117 residues exists within P450R between the FAD and FAD/NADPH domains (also see Porter, 1991). This insertion domain is thought to optimise the relative positions of the FMN- and FAD-binding domains in order to permit efficient electron transfer between the two flavins, and may also interact with cytochrome P_{450} (Porter *et al.*, 1986).

THREADER (Jones *et al.*, 1992b) identifies rat *oncomodulin* (calcium binding protein; PDB code 1RRO) as the most probable reference for the fold of this insertion domain. The significance score for the filtered pairwise energies is -2.97, which is categorised by THREADER as 'borderline significant' (carbonmonoxy erythrocruorin or haemoglobin, PDB code 1ECO, had the next highest score of -2.67, which is categorised as a 'poor score'). The significance score for the weighted sum of pairwise and solvation energy is -2.37 (a guideline score for significance in this case is less than -2.0).

TOPITS (Rost, 1995a,b) assigned bacterial aspartate receptor-ligand binding domain (PDB code 1WAS). The significance score, *ZALI*, is 1.75 which is categorised by TOPITS as having a 33% chance of being correct (β-amylase, PDB code 1TML, had the next highest score of 1.74, which is also categorised as having a 33% chance of being correct).

A probable fold from one of these methods was not rated as a probable fold by the other method. For reasons of endpoint geometry (see below), the 3D structure of the insertion domain was modelled on 1RRO using the THREADER output alignment (Figure 3.4). Comparison of the secondary structure of the identified fold and the secondary structure prediction of this region of P450R using PHDsec (Rost & Sander, 1993a; Rost & Sander, 1993b; Rost & Sander, 1994a; Rost *et al.*, 1994) did not give significant justification for refinement of any regions of the alignment. The absence of consistent structural violations across the ensemble indicated an acceptable sequence alignment also.



```
P450R Sec Hom
P450R Sec X-ray
P450R Res. No.   330                 340                 350
P450R            L V N Q L G K I L G A D L D V V M S L N N L D E E S N K
1RRO             - - - S I T D - - I L S A E D I A A A L Q E C Q D P D T

P450R Sec Hom
P450R Sec X-ray
P450R Res. No.       360                 370                 380
P450R            K H P F P C P T S Y R T A L T Y Y L D I T N P P R T N V
1RRO             - - - - - - - F E P Q K F F Q T S G - L S K M - S A S Q

P450R Sec Hom
P450R Sec X-ray
P450R Res. No.           390                       400
P450R            L Y E L A Q Y A S E P S E - - - - Q E L L R K M A S S S
1RRO             V K D I F R F I D N D Q S G Y L D G D E L K Y F L Q K F

P450R Sec Hom
P450R Sec X-ray
P450R Res. No.   410                 420                 430
P450R            G E G K E L Y L S W V V E A R R H I L A I L Q D C P S -
1RRO             Q S D - - - A R E L T E S E T K S L M D A A D N D G D G

P450R Sec Hom
P450R Sec X-ray
P450R Res. No.           440
P450R            - - L R P P I D H L C E -
1RRO             K I G A D E F Q E M V H S
```

Figure 3.4 Threading-based alignment for the P450R 'insertion domain'. Row 1 shows the secondary structure of the predicted model. Row 2 shows the secondary structure of the corresponding residues in the crystal structure of Wang *et al.* (1997).

Although the secondary structures in the predicted model and the crystal structure are largely α-helical, the structural topologies are somewhat different between the two.

The insertion domain was modelled simultaneously with the FAD and FAD/NADPH regions producing a model comprising of a single polypeptide chain. The C-terminus of the FAD domain was in close spatial proximity to the N-terminus of the FAD/NADPH domain (see section 3.3.2.2). This required that the N- and C-termini of the insertion domain would have to be in relatively close proximity. The THREADER result, 1RRO, possesses a more suitable N- and C-terminal spatial separation for this purpose than the PHDthreader result, 1WAS, which had a large endpoint separation (*i.e.* endpoint Cα-Cα distance of 31Å).

For the representative model of the FAD + FAD/NADPH and insertion domains, the overall G-factor, as determined by PROCHECK (Laskowski *et al.*, 1993), was 0.00 (ideally G-factor scores should be above -0.5; values below -1.0 may need investigation), and 85% of residues in this structure had their φ/ψ angles in the most favoured regions of Ramachandran space. This signifies satisfactory quality of stereochemical structure. However, while the subsequently-derived crystal structure solved by Wang *et al.* (1997) shows that this domain is largely α-helical, the locations of the helices in the model are, in places, significantly different to those in the crystal structure (see Figure 3.4). Also, a helix is predicted in the model where a β-strand exists in the crystal structure (see Figure 3.4).

3.3.2 Homology-Modelled Domains of P450R

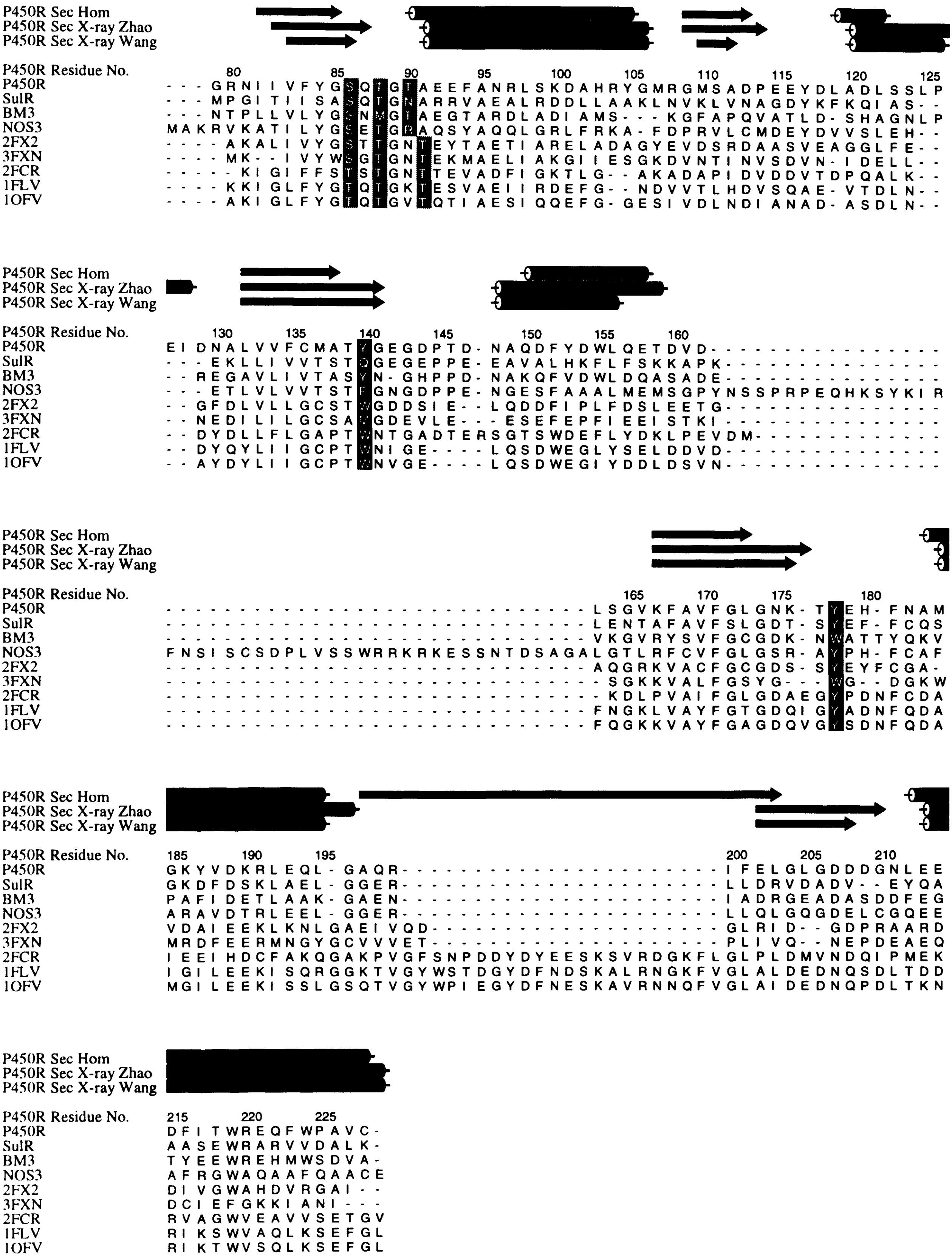
The domains which were modelled based on sequence homology are the FMN, FAD and FAD/NADPH domains (refer to Figure 3.1).

3.3.2.1 FMN Domain

The protein with known 3D structure with highest sequence homology with the FMN domain of P450R was found by BLAST to be *Desulfovibrio vulgaris* flavodoxin, with *p(N)* of approximately 10^{-5}. Flavodoxins (all bacterial) found with known 3D structure were as follows: *Desulfovibrio vulgaris* (PDB code 2FX2; Watt *et al.*, 1991), *Clostridium MP* (PDB code 3FXN; Smith *et al.*, 1977), *Chondrus crispus* (PDB code 2FCR; Fukuyama *et al.*, 1992), *Anabaena 7120* (PDB code 1FLV; Rao *et al.*, 1992) and *Anacystis nidulans* (PDB code 1OFV; Ludwig *et al.*, 1992) flavodoxins. The 3D structures of the flavodoxins were used as templates for the P450R model.

The multiple sequence alignment of P450R, BM3, SulR, NOS3 and the five bacterial flavodoxins is given in Figure 3.5. The interior and exterior flavin ring shielding residues (P450R: Tyr-140 and Tyr-178, respectively) as well as the phosphate binding residues (P450R: Ser-86, Thr-88 and Thr-90) are highlighted. The alignment also compares the predicted secondary structure of P450R, the secondary structure of P450R taken directly from the model, and the secondary structure derived from NMR data performed by the collaborators (available subsequent to model building).

Three of the flavodoxins possess a loop of approximately 20 residues in length near the C-terminus which does not exist in the other proteins in the alignment (Figures 3.5 and 3.6). The lack of this loop within P450R was found to be important in the docking of the flavin domains (see section 3.3.3), where its absence permitted a reasonable proximity between the flavin prosthetic groups.



```
P450R Sec Hom
P450R Sec X-ray Zhao
P450R Sec X-ray Wang

P450R Residue No.      80     85   90   95   100  105  110  115  120  125
P450R               ---GRNIIVFYGSQTGTAEEFANRLSKDAHRYGMRGMSADPEEYDLADLSSLP
SulR                ---MPGITIISASQTGNARRVAEALRDDLLAAKLNVKLVNAGDYKFKQIAS--
BM3                 ---NTPLLVLYGSNMGTAEGTARDLADIAMS---KGFAPQVATLD-SHAGNLP
NOS3                MAKRVKATILYGSETGRAQSYAQQLGRLFRKA-FDPRVLCMDEYDVVSLEH--
2FX2                ----AKALIVYGSTTGNTEYTAETIARELADAGYEVDSRDAASVEAGGLFE--
3FXN                ----MK--IVYWSGTGNTEKMAELIAKGIIESGKDVNTINVSDVN-IDELL--
2FCR                -----KIGIFFSTSTGNTTEVADFIGKTLG--AKADAPIDVDDVTDPQALK--
1FLV                ----KKIGLFYGTQTGKTESVAEIIRDEFG--NDVVTLHDVSQAE-VTDLN--
1OFV                ----AKIGLFYGTQTGVTQTIAESIQQEFG-GESIVDLNDIANAD-ASDLN--

P450R Sec Hom
P450R Sec X-ray Zhao
P450R Sec X-ray Wang

P450R Residue No.      130  135  140  145  150  155  160
P450R               EIDNALVVFCMATYGEGDPTD-NAQDFYDWLQETDVD----------------
SulR                ---EKLLIVVTSTQGEGEPPE-EAVALHKFLFSKKAPK---------------
BM3                 --REGAVLIVTASYN-GHPPD-NAKQFVDWLDQASADE---------------
NOS3                ---ETLVLVVTSTFGNGDPPE-NGESFAAALMEMSGPYNSSPRPEQHKSYKIR
2FX2                --GFDLVLLGCSTWGDDSIE--LQDDFIPLFDSLEETG---------------
3FXN                --NEDILILGCSAYGDEVLE---ESEFEPFIEEISTKI---------------
2FCR                --DYDLLFLGAPTWNTGADTERSGTSWDEFLYDKLPEVDM-------------
1FLV                --DYQYLIIGCPTWNIGE----LQSDWEGLYSELDDVD---------------
1OFV                --AYDYLIIGCPTWNVGE----LQSDWEGIYDDLDSVN---------------

P450R Sec Hom
P450R Sec X-ray Zhao
P450R Sec X-ray Wang

P450R Residue No.                                  165  170  175   180
P450R               -----------------------------LSGVKFAVFGLGNK-TYEH-FNAM
SulR                -----------------------------LENTAFAVFSLGDT-SYEF-FCQS
BM3                 -----------------------------VKGVRYSVFGCGDK-NWATTYQKV
NOS3                FNSISCSDPLVSSWRRKRKESSNTDSAGALGTLRFCVFGLGSR-AYPH-FCAF
2FX2                -----------------------------AQGRKVACFGCGDS-SYEYFCGA-
3FXN                ------------------------------SGKKVALFGSYG---WG--DGKW
2FCR                ------------------------------KDLPVAIFGLGDAEGYPDNFCDA
1FLV                -----------------------------FNGKLVAYFGTGDQIGYADNFQDA
1OFV                -----------------------------FQGKKVAYFGAGDQVGYSDNFQDA

P450R Sec Hom
P450R Sec X-ray Zhao
P450R Sec X-ray Wang

P450R Residue No.   185  190  195                          200  205  210
P450R               GKYVDKRLEQL-GAQR----------------------IFELGLGDDDGNLEE
SulR                GKDFDSKLAEL-GGER----------------------LLDRVDADV--EYQA
BM3                 PAFIDETLAAK-GAEN----------------------IADRGEADASDDFEG
NOS3                ARAVDTRLEEL-GGER----------------------LLQLGQGDELCGQEE
2FX2                VDAIEEKLKNLGAEIVQD--------------------GLRID--GDPRAARD
3FXN                MRDFEERMNGYGCVVVET--------------------PLIVQ--NEPDEAEQ
2FCR                IEEIHDCFAKQGAKPVGFSNPDDYDYEESKSVRDGKFLGLPLDMVNDQIPMEK
1FLV                IGILEEKISQRGGKTVGYWSTDGYDFNDSKALRNGKFVGLALDEDNQSDLTDD
1OFV                MGILEEKISSLGSQTVGYWPIEGYDFNESKAVRNNQFVGLAIDEDNQPDLTKN

P450R Sec Hom
P450R Sec X-ray Zhao
P450R Sec X-ray Wang

P450R Residue No.   215  220  225
P450R               DFITWREQFWPAVC-
SulR                AASEWRARVVDALK-
BM3                 TYEEWREHMWSDVA-
NOS3                AFRGWAQAAFQAACE
2FX2                DIVGWAHDVRGAI--
3FXN                DCIEFGKKIANI---
2FCR                RVAGWVEAVVSETGV
1FLV                RIKSWVAQLKSEFGL
1OFV                RIKTWVSQLKSEFGL
```

Figure 3.5 Sequence alignment of P450R FMN domain. Row 1 shows the secondary structure of the homology model. Row 2 shows the secondary structure of the FMN domain crystal structure of Zhao *et al.* (1996). Row 3 shows the secondary structure of corresponding residues in the crystal structure of Wang *et al.* (1997). Flavin ring-shielding residues are shown on a black background, and phosphate binding residues are shown on a grey background.

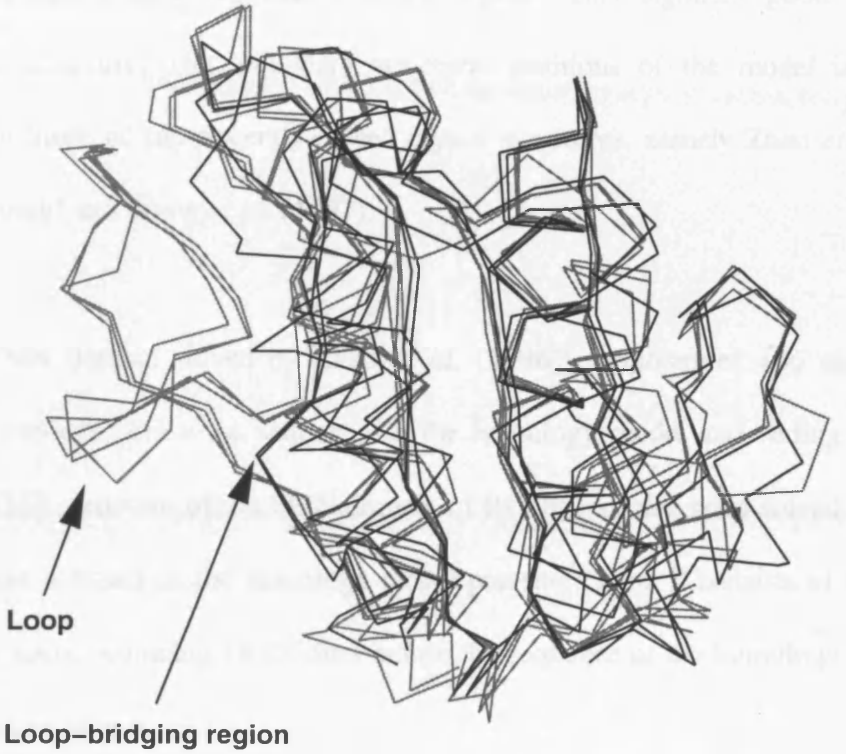


Figure 3.6 Cα superposition of five bacterial flavodoxin crystal structures. A loop, displayed towards the left, exists in three of these structures (light) and is bridged in the case of the other two structures (dark).

A representation of the model for the FMN domain is given in Figure 3.7.

The overall G-factor, as determined by PROCHECK (Laskowski *et al.*, 1993), for the representative model was 0.06 (ideally G-factor scores should be above -0.5; values below -1.0 may need investigation), and 92% of residues in this structure had their ϕ/ψ angles in the most favoured regions of Ramachandran space. This signifies good quality of stereochemical structure. The secondary structural positions of the model is relatively consistent with those of subsequently-solved crystal structures, namely Zhao *et al.* (1996; FMN domain only) and Wang *et al.* (1997).

The P450R FMN domain solved by Zhao *et al.* (1996) comprises of 166 amino acids, beginning 10 residues before the sequence of the homology model and ending 4 residues after it. The NMR structure of the FMN domain of P450R has also been solved (Barsukov *et al.*, 1997) and is based on the homology model presented here; it consists of assignment for 179 amino acids, beginning 14 residues before the sequence of the homology model and ending 13 residues after it.

A significant difference regarding overall fold between the crystal structure of the FMN domain of P450R and flavodoxin structures is that the helices flanking the β-sheet are shifted significantly when the β-sheets are superposed.

In the flavodoxin structures, a β-bulge is present in strand β5 which is a site where some flavodoxins contain an insertion sequence (insertion site illustrated in Figure 3.6). Such a β-bulge also occurs in the P450R FMN domain crystal structure. However, the homology model was built with an unbroken β-strand since this was predicted to be the case when

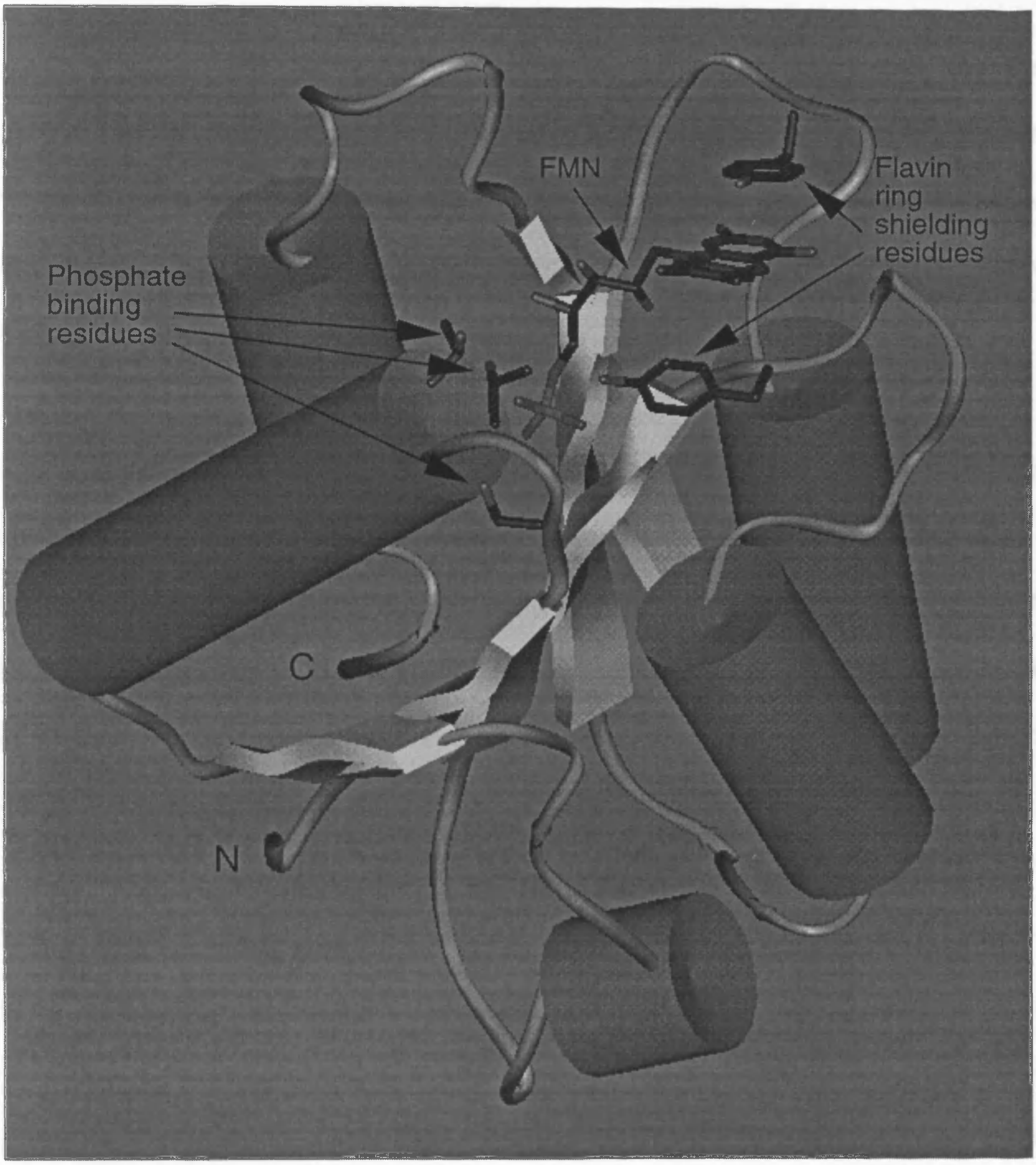


Figure 3.7 Homology model for FMN binding domain of P450R. Strands are represented by ribbons; helices are represented by cylinders.

studying the structure-based sequence alignment of this domain. Whether or not the NMR structure contains this β-bulge is not specified by Barsukov *et al.* (1997).

A helical region is reported in both the crystal and NMR structures after strand β2 in the FMN domain (this region is specified as consisting of two separate 3_{10} helices in the case of the crystal structure), whereas this exists as a loop region in the flavodoxin structures. A helix-like conformation also occurs within the homology model of the FMN domain for this region.

A helix, specified in Zhao *et al.* (1996) as an α-helix, is also reported in both experimentally determined structures to exist in the residues preceding the sequences of the homology-defined FMN domain. In the crystal structure, this helix is shown to pack against the rest of the FMN domain, whereas it is reported to be not well-defined in the NMR structure. A helical structure was not predicted for these residues, which were modelled as part of the membrane binding domain (Section 3.3.1.1).

The four residues in the crystal structure FMN domain that continue beyond the homology-defined FMN domain are shown to extend the C-terminal α-helix. This is consistent with the NMR structure, where a one-turn extension on the Porter & Kasper (1986) homology-defined FMN domain (consistent with the homology-defined FMN domain presented in this study) is reported on the C-terminal helix. The remainder of the NMR structure C-terminus seems to be not well-defined.

Barsukov *et al.* (1997) found limited chemical shift changes between the oxidised and reduced states of FMN, suggesting a relatively limited change in conformation in the FMN domain between these oxidation states; the four residues that are affected are spatially remote from the isoalloxazine ring.

The Cα superposition of the 3D structures of the P450R model and the crystal structure for the FMN domain is shown in Figure 3.8. The Cα RMSD between the crystal structure and the homology model over all common atoms is 3.9Å (over the β-sheet alone the value is 0.75Å, and over the α-helices alone 1.99Å).

3.3.2.2 FAD and FAD/NADPH Domains

The FAD and FAD/NADPH domains together were found to exhibit highest sequence homology with spinach ferredoxin $NADP^+$ reductase, with a BLAST *p(N)* of approximately 10^{-12}. Ferredoxin reductase-like proteins found with known structure were as follows: spinach ferredoxin $NADP^+$ reductase (FNR; PDB code 1FNR, Karplus *et al.*, 1991), corn nitrate reductase (NR; PDB code 1CND, Lu *et al.*, 1994) and pig NADH-cytochrome b5 reductase (b5R; PDB code 1NDH, Nishida *et al.*, to be published), all of which are FAD containing proteins. FNR is NADPH binding, whereas NR and b5R are NADH binding proteins.

The crystal structure for the NADPH analogue, 2'-phospho-AMP (2'-phospho-adenosine monophosphate), bound to spinach FNR (PDB code 2FNR, Karplus *et al.*, 1991) shows that the FAD/NADPH domain is indeed expected to be the site for NADPH binding, as well as for FAD binding, where the FAD and NADPH bind at different but adjacent positions on the protein.

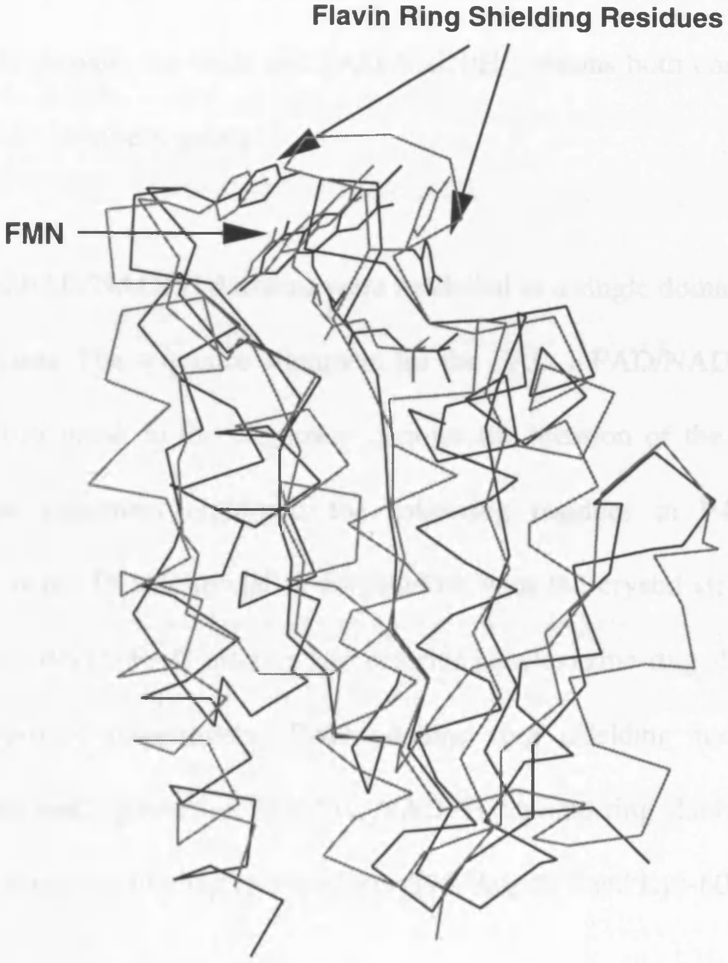
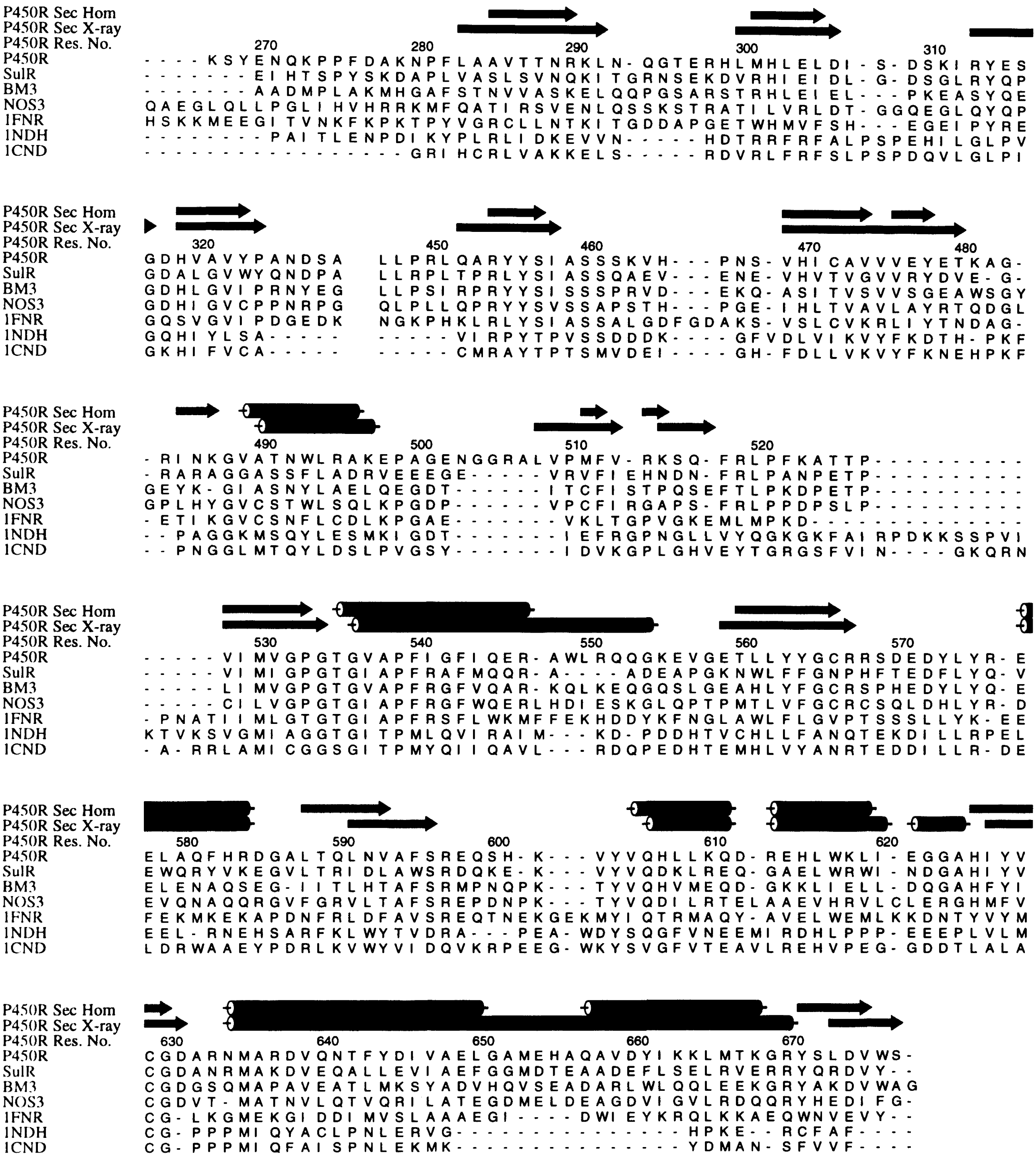


Figure 3.8 Cα superposition of P450R crystal structure of FMN domain (Zhao *et al.*, 1996) and homology model of FMN domain (dark and light, respectively) over common atoms.

Although the P450R FAD and FAD/NADPH domains are interrupted by a 117 residue insertion sequence (see section 3.3.1.3), the FAD domain forms an integral part of the FAD/NADPH domain - the FAD domain is intertwined into part of the FAD/NADPH domain. For example, based on the structures of the homology-identified proteins, part of the FAD domain comprises of a strand of a β-sheet, of which the rest is formed by the FAD/NADPH domain; the FAD and FAD/NADPH domains both contribute to the binding site of the FAD prosthetic group.

The FAD and FAD/NADPH domains were modelled as a single domain using FNR, NR and b5R as templates. The sequence alignment for the FAD + FAD/NADPH domains is shown in Figure 3.9 (a break in the alignment denotes the location of the 117 residue insertion domain). The alignment highlights the following residues in P450R (determined by examination of the P450R model in conjunction with the crystal structure study of FNR, Karplus *et al.*, 1991): FAD interior and exterior isoalloxazine ring shielding residues, Tyr-456 and Trp-676, respectively; FAD adenine ring shielding residue, Tyr-478; FAD pyrophosphate binding residue, Arg-454; NADPH adenine ring shielding residue, Tyr-604; NADPH 2'-phosphate binding residues, Ser-596, Arg-597 and Lys-602.

In their study of FNR, Karplus *et al.* (1991) also identify the residue equivalent to P450R Cys-629 as a putative NADPH nicotinamide binding residue. Additionally, the nicotinamide group in NADPH is thought to interact with FAD by stacking onto the flavin ring in the FAD prosthetic group (Karplus *et al.*, 1991), suggesting direct electron transfer from NADPH to the FAD prosthetic group in P450R.



```
P450R Sec Hom
P450R Sec X-ray
P450R Res. No.            270       280       290       300       310
P450R            ----KSYENQKPPFDAKNPFLAAVTTNRKLN-QGTERHLMHLELDI-S-DSKIRYES
SulR             -------EIHTSPYSKDAPLVASLSVNQKITGRNSEKDVRHIEIDL-G-DSGLRYQP
BM3              -------AADMPLAKMHGAFSTNVVASKELQQPGSARSTRHLEIEL---PKEASYQE
NOS3             QAEGLQLLPGLIHVHRRKMFQATIRSVENLQSSKSTRATILVRLDT-GGQEGLQYQP
1FNR             HSKKMEEGITVNKFKPKTPYVGRCLLNTKITGDDAPGETWHMVFSH---EGEIPYRE
1NDH             --------PAITLENPDIKYPLRLIDKEVVN-----HDTRRFRFALPSPEHILGLPV
1CND             -----------------GRIHCRLVAKKELS-----RDVRLFRFSLPSPDQVLGLPI

P450R Sec Hom
P450R Sec X-ray
P450R Res. No.     320              450       460        470       480
P450R            GDHVAVYPANDSA   LLPRLQARYYSIASSSKVH---PNS-VHICAVVVEYETKAG-
SulR             GDALGVWYQNDPA   LLRPLTPRLYSIASSQAEV---ENE-VHVTVGVVRYDVE-G-
BM3              GDHLGVIPRNYEG   LLPSIRPRYYSISSSPRVD---EKQ-ASITVSVVSGEAWSGY
NOS3             GDHIGVCPPNRPG   QLPLLQPRYYSVSSAPSTH---PGE-IHLTVAVLAYRTQDGL
1FNR             GQSVGVIPDGEDK   NGKPHKLRLYSIASSALGDFGDAKS-VSLCVKRLIYTNDAG-
1NDH             GQHIYLSA-----   -----VIRPYTPVSSDDDK----GFVDLVIKVYFKDTH-PKF
1CND             GKHIFVCA-----   -----CMRAYTPTSMVDEI----GH-FDLLVKVYFKNEHPKF

P450R Sec Hom
P450R Sec X-ray
P450R Res. No.           490       500       510        520
P450R            -RINKGVATNWLRAKEPAGENGGRALVPMFV-RKSQ-FRLPFKATTP----------
SulR             -RARAGGASSFLADRVEEEGE-----VRVFIEHNDN-FRLPANPETP----------
BM3              GEYK-GIASNYLAELQEGDT------ITCFISTPQSEFTLPKDPETP----------
NOS3             GPLHYGVCSTWLSQLKPGDP------VPCFIRGAPS-FRLPPDPSLP----------
1FNR             -ETIKGVCSNFLCDLKPGAE-------VKLTGPVGKEMLMPKD--------------
1NDH             --PAGGKMSQYLESMKIGDT-------IEFRGPNGLLVYQGKGKFAIRPDKKSSPVI
1CND             --PNGGLMTQYLDSLPVGSY-------IDVKGPLGHVEYTGRGSFVIN----GKQRN

P450R Sec Hom
P450R Sec X-ray
P450R Res. No.            530       540        550       560       570
P450R            -----VIMVGPGTGVAPFIGFIQER-AWLRQQGKEVGETLLYYGCRRSDEDYLYR-E
SulR             -----VIMIGPGTGIAPFRAFMQQR-A----ADEAPGKNWLFFGNPHFTEDFLYQ-V
BM3              -----LIMVGPGTGVAPFRGFVQAR-KQLKEQGQSLGEAHLYFGCRSPHEDYLYQ-E
NOS3             -----CILVGPGTGIAPFRGFWQERLHDIESKGLQPTPMTLVFGCRCSQLDHLYR-D
1FNR             -PNATIIMLGTGTGIAPFRSFLWKMFFEKHDDYKFNGLAWLFLGVPTSSSLLYK-EE
1NDH             KTVKSVGMIAGGTGITPMLQVIRAIM---KD-PDDHTVCHLLFANQTEKDILLRPEL
1CND             -A-RRLAMICGGSGITPMYQIIQAVL---RDQPEDHTEMHLVYANRTEDDILLR-DE

P450R Sec Hom
P450R Sec X-ray
P450R Res. No.     580       590        600       610       620
P450R            ELAQFHRDGALTQLNVAFSREQSH-K---VYVQHLLKQD-REHLWKLI-EGGAHIYV
SulR             EWQRYVKEGVLTRIDLAWSRDQKE-K---VYVQDKLREQ-GAELWRWI-NDGAHIYV
BM3              ELENAQSEG-IITLHTAFSRMPNQPK---TYVQHVMEQD-GKKLIELL-DQGAHFYI
NOS3             EVQNAQQRGVFGRVLTAFSREPDNPK---TYVQDILRTELAAEVHRVLCLERGHMFV
1FNR             FEKMKEKAPDNFRLDFAVSREQTNEKGEKMYIQTRMAQY-AVELWEMLKKDNTYVYM
1NDH             EEL-RNEHSARFKLWYTVDRA---PEA-WDYSQGFVNEEMIRDHLPPP-EEEPLVLM
1CND             LDRWAAEYPDRLKVWYVIDQVKRPEEG-WKYSVGFVTEAVLREHVPEG-GDDTLALA

P450R Sec Hom
P450R Sec X-ray
P450R Res. No.   630       640       650       660       670
P450R            CGDARNMARDVQNTFYDIVAELGAMEHAQAVDYIKKLMTKGRYSLDVWS-
SulR             CGDANRMAKDVEQALLEVIAEFGGMDTEAADEFLSELRVERRYQRDVY--
BM3              CGDGSQMAPAVEATLMKSYADVHQVSEADARLWLQQLEEKGRYAKDVWAG
NOS3             CGDVT-MATNVLQTVQRILATEGDMELDEAGDVIGVLRDQQRYHEDIFG-
1FNR             CG-LKGMEKGIDDIMVSLAAAEGI----DWIEYKRQLKKAEQWNVEVY--
1NDH             CG-PPPMIQYACLPNLERVG----------------HPKE--RCFAF----
1CND             CG-PPPMIQFAISPNLEKMK----------------YDMAN-SFVVF----
```

Figure 3.9 Sequence alignment of P450R 'FAD + FAD/NADPH' domain. The break in the sequences indicates the location of the 'insertion domain' in P450R. Row 1 shows the secondary structure of the homology model. Row 2 shows the secondary structure of thecorresponding residues in crystal structure of Wang *et al.* (1997). In the crystal structure, residues 634-670 were assigned as two separate helices; however, the residue numbers that separate them were not specified.

The 117 residue insertion domain (see section 3.3.1.3) in P450R exists at the location of a protruding loop that originates from, and reforms into, the FAD + FAD/NADPH domain. The insertion domain therefore exists in an easily-defined spatial position with respect to the FAD + FAD/NADPH domain. It was possible to manually dock the template structures for the FAD + FAD/NADPH and 'insertion' domains into a 'reasonable' relative position before automatic model generation. The models for the FAD, 'insertion' and FAD/NADPH domains were then able to be generated simultaneously. Figure 3.10 shows the relative positioning of the FAD + FAD/NADPH and insertion domains.

To avoid MODELLER structural violations, two regions of the NR and b5R template structures were removed, thus focusing only on FNR in these regions. One of these regions corresponds to a loop that exists only in the NADH binding proteins. The other was a significant difference in backbone conformation between FNR and the NADH binding proteins, relatively close to the NAD(P)H binding site (*i.e.* equivalent to the 2'-phospho-AMP binding site in the 2FNR crystal structure). The biasing of the P450R model towards FNR in these regions is justified on the basis that (1) P450R shares higher sequence similarity with FNR than with NR or b5R, and (2) P450R and FNR are NADPH-specific.

For the representative model of the FAD + FAD/NADPH and insertion domains, the overall G-factor, as determined by PROCHECK (Laskowski *et al.*, 1993), was 0.00 (ideally G-factor scores should be above -0.5; values below -1.0 may need investigation), and 85% of residues in this structure had their ϕ/ψ angles in the most favoured regions of Ramachandran space. This signifies satisfactory quality of stereochemical structure.

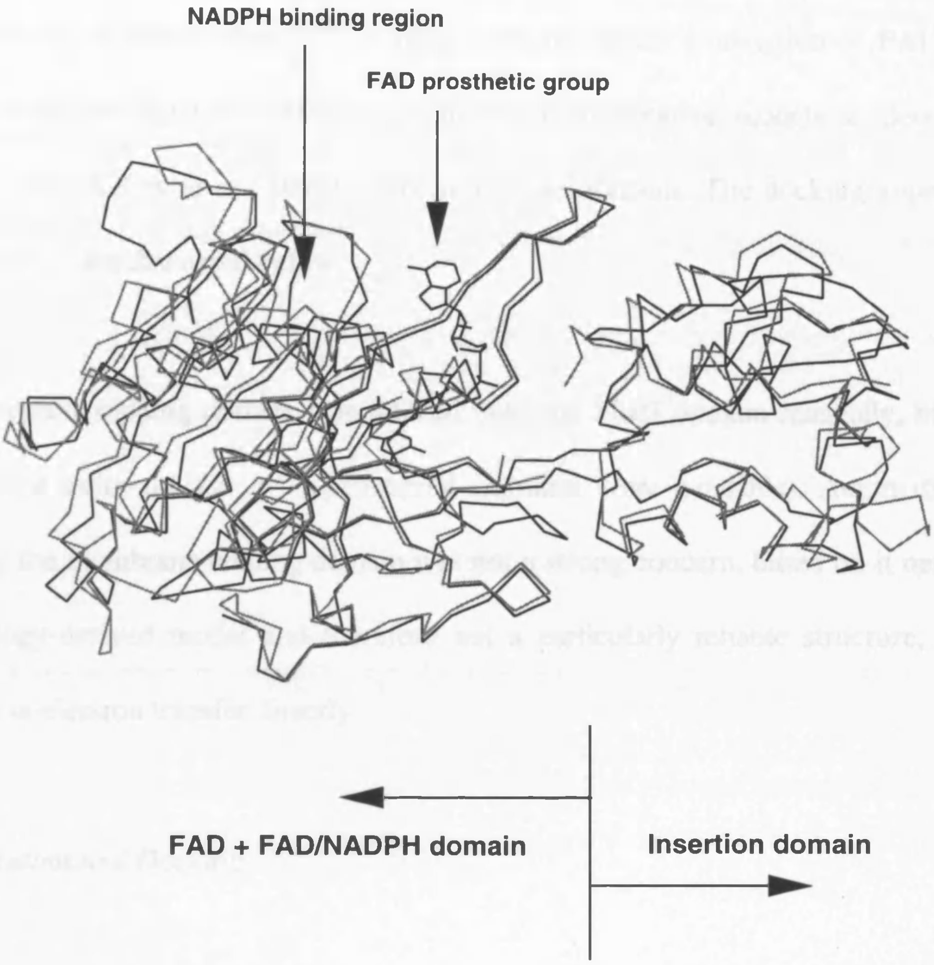


Figure 3.10 Cα trace of P450R FAD+FAD/NADPH and 'insertion' domains (dark) superimposed on FNR and oncomodulin crystal structures, respectively (light).

3.3.3 Docking of the Separate Domains

The inter-domain docking for P450R focused on the flavin-containing domains, which are homology-based, rather than threading-based, models. The docking is thus concerned with the interaction between the FMN domain and the 'FAD + insertion + FAD/NADPH' domain. The docking was performed using the representative models as determined by NMRCLUST (Kelley *et al.*, 1996) for both of these domains. The docking approaches for these domains are discussed below.

The membrane binding domain was docked onto the FMN domain manually, but with no well-defined rationale (*i.e.* no experimental restraints were available). Automatic docking involving the membrane binding domain was not a strong concern, based on it neither being a homology-derived model and therefore not a particularly reliable structure, nor being involved in electron transfer directly.

3.3.3.1 Automated Docking

An automated approach to docking was attempted using DOCK (Shoichet & Kuntz, 1991). Both the domains to be docked were relatively large; therefore, the spheres that represent each of the two domains were clustered into suitably sized sub-populations. Each sub-cluster from one of the two domains was docked onto each sub-cluster of the other. Also, both domains were treated, in turn, as 'receptor' and 'ligand'. This resulted in a total of 71 permutations for docking.

All the docking permutations were performed for geometry-based docking. The lowest overall energy output for the docking may not necessarily be the 'correct' result. Such 'false positives' were filtered out using a flavin-flavin distance check of 17Å subsequent to the docking of each permutation, since it was reported (unpublished results) that the flavin-flavin distance in P450R was found to be within a distance of 14Å (an error of 3Å, *i.e.* approximately 20%, was used arbitrarily for the purpose of the distance check).

None of the automated docking runs produced domains where the flavin groups were within the 17Å cutoff distance. It was therefore not necessary to perform the forcefield docking (used for docking refinement) since all the geometry-based docking models were discarded.

These automated docking runs were not sucessful due to the fact that docking programs are based on docking a 'ligand' into a 'receptor'. However, larger molecules, such as protein domains, may interact based on steric and chemical complementarity over a relatively large surface area and a relatively flat interface.

3.3.3.2 Manual Docking

Electrostatic potentials for both the FMN domain and the 'FAD + insertion + FAD/NADPH' domain were calculated using DELPHI (Gilson *et al.*, 1987; see also Nicholls & Honig, 1991). It was found using GRASP (Nicholls *et al.*, 1991), that both these domains possessed dipole moment vectors with significant magnitude; the magnitude of the dipole for the FMN domain was 514 Debye (for all of the FMN domain crystal structure the dipole moment was 677 Debye). The dipole moment vector of the homology-modelled

FMN domain is in a very similar direction to that of the crystal structure. The dipole moment for the model of the 'FAD + insertion + FAD/NADPH' domain was 756 Debye. The dipole moment vectors, along with consideration of surface charge complementarity between the two domains as well as steric considerations, were used as guides to orientate and dock the domains (Figure 3.11). The flavin-flavin distance was not taken into account in this procedure.

The spatial alignment of dipoles was justified by the fact that a molecular dipole has been described in cytochrome P_{450} BM3 suggesting that the dipole might aid in redox partner docking (Hasemann *et al.*, 1995). GRASP was used to calculate its dipole moment vector, giving a magnitude of 698 Debye. The dipole moment could act as an initial inter-domain positioning factor; however, the interaction between the FMN and FAD redox centres is expected to be predominant at a range where electron transfer is able to take place.

The resulting manually docked domains (Figure 3.12) exhibit an extended β-sheet / β-sheet positioning between the FMN and FAD/NADPH domain. In the docked model, the FMN and FAD prosthetic groups are located 13.7Å apart, with the isoalloxazine rings oriented toward each other. Such proximity of the flavins is made possible (using the above manual docking criteria) because of the absence of a loop in the FMN domain of P450R which is present as a sequence approximately 20 residues in length in some of the template flavodoxins. The absence of this loop in P450R also produced a relatively flat surface in the FMN domain which could be docked onto a relatively flat surface in the FAD/NADPH domain.

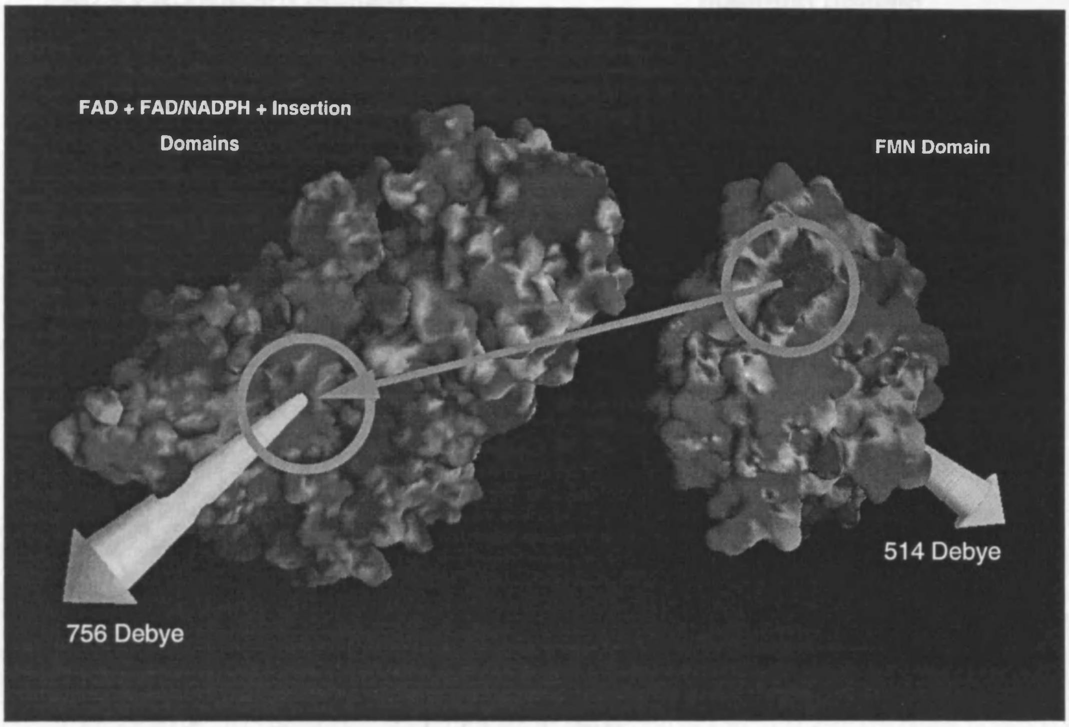


Figure 3.11 Electrostatic surface representation of separate P450R domains illustrating opposing faces of electrostatic surface potential. The circles and the arrow between them indicate complementary regions whereby a positive region on the FMN domain is to be fitted onto a negative region on the 'FAD + FAD/NADPH + Insertion' domain in such a way that the dipole moment vectors (remaining two arrows) become aligned.

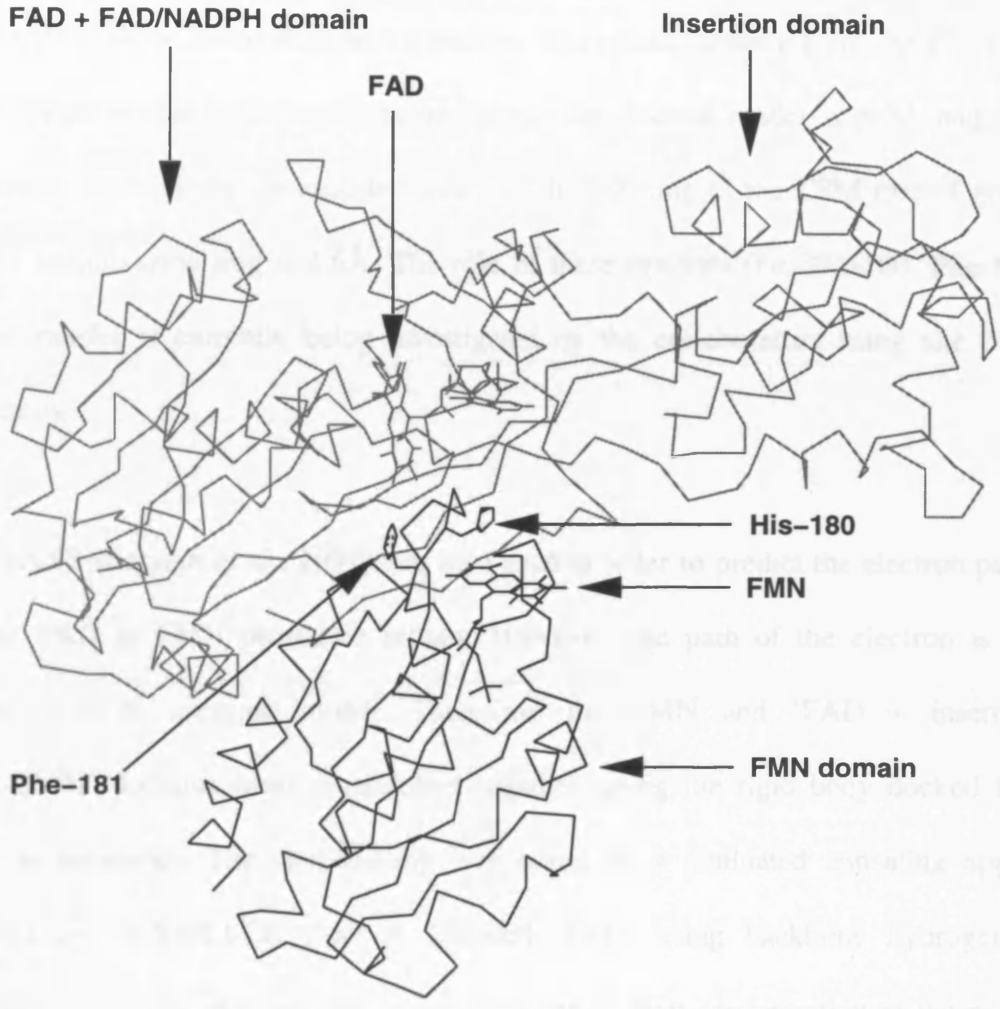


Figure 3.12 Docked domains for P450R model: FMN domain (dark) and FAD + FAD/NADPH + Insertion domains (light). The two residues (His–180 and Phe–181) identified as possibly being involved in the electron transfer mechanism are illustrated.

Manual inspection of the docked model's 3D structure identifies two residues, both in the FMN domain, that may be involved in electron transfer from the FAD to the FMN prosthetic groups, namely, His-180 and Phe-181. The sequence alignment of the FMN domain (Figure 3.5) adds weight to this prediction based on the relative similarities in residue type at the relevant alignment positions. The spatial distance from the His-180 and Phe-181 rings to the FAD isoalloxazine ring in the docked model is 6.9Å and 10.4Å, respectively, although the distance between the His-180 ring to the C8M methyl group of the FAD isoalloxazine ring is 4.6Å. The role of these residues (*i.e.* His-180, Phe-181) in electron transfer is currently being investigated by the collaborators using site directed mutagenesis.

PATHWAYS (Beratan *et al.*, 1991) was also used in order to predict the electron pathway from the FAD to FMN prosthetic groups. However, the path of the electron is highly dependent on an accurate model. Therefore, the FMN and 'FAD + insertion + FAD/NADPH' domains were re-modelled together, using the rigid body docked P450R models as templates. The re-modelling was based on a simluated annealing approach performed by MODELLER (Sali & Blundell, 1993) using backbone hydrogen-bond restraints between the adjacent ends of the extended β-sheet conformation of the manually docked model (see Table 3.1). The most representative model of the ensemble, as determined by NMRCLUST (Kelley *et al.*, 1996), was used for the PATHWAYS calculation.

The re-modelling produced an FMN-FAD prosthetic group distance of 9.4Å, and a distance between the His-180 and Phe-181 rings to the FAD isoalloxazine ring of 6.8Å and 9.4Å, respectively. However, PATHWAYS predicted that the electron transfer from FAD to

FMN is mediated solely by the His-180 backbone nitrogen; this may suggest that the fine details of the models may not be highly accurate, but it may also suggest a somewhat simplistic treatment of electronic orbitals.

β-strand in FMN domain	β-strand in FAD+FAD/NADPH domain	Upper Bound (Å)	Lower Bound (Å)
Glu-202 N	Tyr-671 O	3.3	2.5
Glu-202 O	Leu-673 N	3.3	2.5
Gly-128 N	Leu-673 O	3.3	2.5
Gly-128 O	Val-675 N	3.3	2.5

Table 3.1 Hydrogen bond distance constraints for P450R domain docking.

The subsequently-derived crystal structure of P450R (Wang *et al.*, 1997; coordinates not yet available) is reported to be bowl-shaped, with the flavins lying in the middle of the concave surface; this seems somewhat similar to the model presented here, although the coordinates are required to make any detailed comparison. A closest distance of 3.5Å is observed between the C7M atoms of the FMN and FAD rings in the crystal structure, and the electron is expected to transfer directly between the flavin groups.

3.4 Conclusion

The P450R model as a whole contained significant regions with insufficient sequence homology with proteins that have 'known' 3D structure (before the P450R structures or the FMN domain were determined experimentally). However, the flavin-containing domains – the domains associated directly with electron transfer – were modelled based on homology and are therefore likely to be modelled more accurately.

The model of the FMN domain, based on homology with bacterial flavodoxins, compares favourably with the crystal structure of this domain (Zhao *et al.*, 1996). However, significant differences exist, such as that related to the global fold where there is a rigid body displacement in the α-helices flanking the β-sheet when P450R is compared with the flavodoxins.

A β-bulge is situated in strand β5 of bacterial flavodoxins, a site where some flavodoxins contain an insertion sequence, and such a β-bulge occurs in the P450R FMN domain crystal structure of Zhao *et al.* (1996). However, the homology model was built with an unbroken β-strand based on examination of the structure-based sequence alignment of the FMN domain.

The Cα RMSD between the crystal structure of the FMN domain of P450R by Zhao *et al.* (1996) and the homology model of this domain over all common atoms is 3.9Å.

The docking of the FMN and 'FAD + insertion + FAD/NADPH' domains using an automated approach was not successful, due to the 'ligand-into-receptor' limitation. A manual approach using dipole moment vectors, electrostatic complementarity and steric complementarity resulted in a flavin-flavin distance of 9.4Å. Two residues, both from the FMN domain of P450R, His-180 and Phe-181, were identified as possibly being involved in electron transfer from the FAD to the FMN prosthetic groups. A crystal structure of P450R appears similar in shape to the docked P450R model. However, a closest distance of 3.5Å is reported between the flavin rings in the crystal structure, and the electron is expected to transfer directly between the flavin groups.

CHAPTER 4

HOMOLOGY MODELLING FOR 'CASP2': THE SECOND CRITICAL ASSESSMENT OF TECHNIQUES FOR PROTEIN STRUCTURE PREDICTION

4.1 Introduction

The first community-wide experiment to assess protein structure prediction methods was CASP1[28] (first Critical Assessment of techniques for protein Structure Prediction), held in 1994 (see Moult *et al.*, 1995). It consisted of three categories: (1) comparative modelling (homology modelling), (2) fold recognition (threading) and (3) *ab initio* folding. 33 protein prediction targets were provided by crystallographers and NMR spectroscopists, 35 research groups took part submitting over 100 predictions. A special issue of PROTEINS: Structure, Function and Genetics (Volume 23, Number 3, 1995) was dedicated to the assessment of CASP1. The assessment of the comparative modelling category was discussed by Mosimann *et al.* (1995).

CASP2[29] (the second Critical Assessment of techniques for protein Structure Prediction) was held in 1996 (see Moult *et al.*, 1997), and included a docking category in addition to the three categories of CASP1. 42 structural targets were provided by experimentalists, 72 research groups participated submitting 947 predictions in total.

[28] URL http://PredictionCenter.llnl.gov/

[29] URL http://PredictionCenter.llnl.gov/

For the CASP2 comparative modelling category, 12 target proteins were provided. The sequence identity of the target protein with the most homologous template protein ranged from 20% to 85%. 19 research groups submitted 81 models containing 122 coordinate sets for assessment. Two members of the Sutcliffe group (Dr Michael J Sutcliffe and Sami Raza), Department of Chemistry, University of Leicester, UK, participated in the comparative modelling category for three of the target proteins, namely Neurocalcin Delta (CASP2 code T0007; modelled by 14 groups in total), Glutathione S-Transferase (CASP2 code T0017; modelled by 11 groups in total) and Ubiquitin Conjugating Enzyme 9 (CASP2 code T0024; modelled by 6 groups in total).

T0007 (neurocalcin delta): Rhodopsin, the photosensitive molecule in retinal rod cells, consists of the protein opsin and the light-sensitive prosthetic group 11-*cis*-retinal. It is a membrane protein containing seven transmembrane helices, with 11-*cis*-retinal lying in the space between the helices. The primary event in visual excitation is the activation of rhodopsin by photon-induced isomerisation of 11-*cis*-retinal to 11-*trans*-retinal at the C11-C12 double bond of this group. Deactivation of rhodopsin occurs by phosphorylation of multiple serine and threonine residues near its C-terminus. Neurocalcin-delta is thought to possibly be involved in the calcium-dependent regulation of rhodopsin phosphorylation.

T0017 (glutathione-S-transferase): The soluble (cytoplasmic) glutathione S-transferases are detoxification enzymes which catalyse the nucleophilic addition of the tri-peptide glutathione (*i.e.* glu-cys-gly) to the electrophilic sites of hydrophobic substrate molecules (both endogenous or xenobiotic). The addition of glutathione increases the solubilities of the substrates, thereby facilitating their excretion from the organism in which they are present.

T0024 (ubiquitin conjugating enzyme 9): Controlled molecular degradation is important for removing abnormal proteins within cells. Such mechanisms exist because, for example, a protein may undergo oxidative damage with the passage of time, or may be synthesised defectively due to errors in translation. When detected in the cell, they are tagged with the protein ubiquitin. In the presence of ubiquitin activating enzyme (E1) and ATP, ubiquitin conjugating enzymes (E2; *e.g.* ubiquitin conjugating enzyme 9) form a thiol-ester adduct with the C-terminal carboxyl group of ubiquitin. Ubiquitin is then transferred from the E2-ubiquitin complex to the target protein in the presence of a target binding protein (E3).

4.2 Methods

The methodological approach to this study is described by the flowchart in Figure 4.1. The methods are detailed in Chapter 2.

Sequence similarity searches were performed using the BLAST algorithm (Altschul *et al.*, 1990), the output of which was used as a basis for 3D modelling.

The SCOP database (Murzin *et al.*, 1995) was used to identify proteins with known structure in the same family as those identified by sequence search. The CATH database (Orengo *et al.*, 1997) exhibited very similar results.

The sequence alignment approach varied between the proteins to be modelled. In the case of T0017, the BLAST alignment was considered sufficiently accurate since the sequence identity between the query sequence and the highest scoring database sequence was 85%

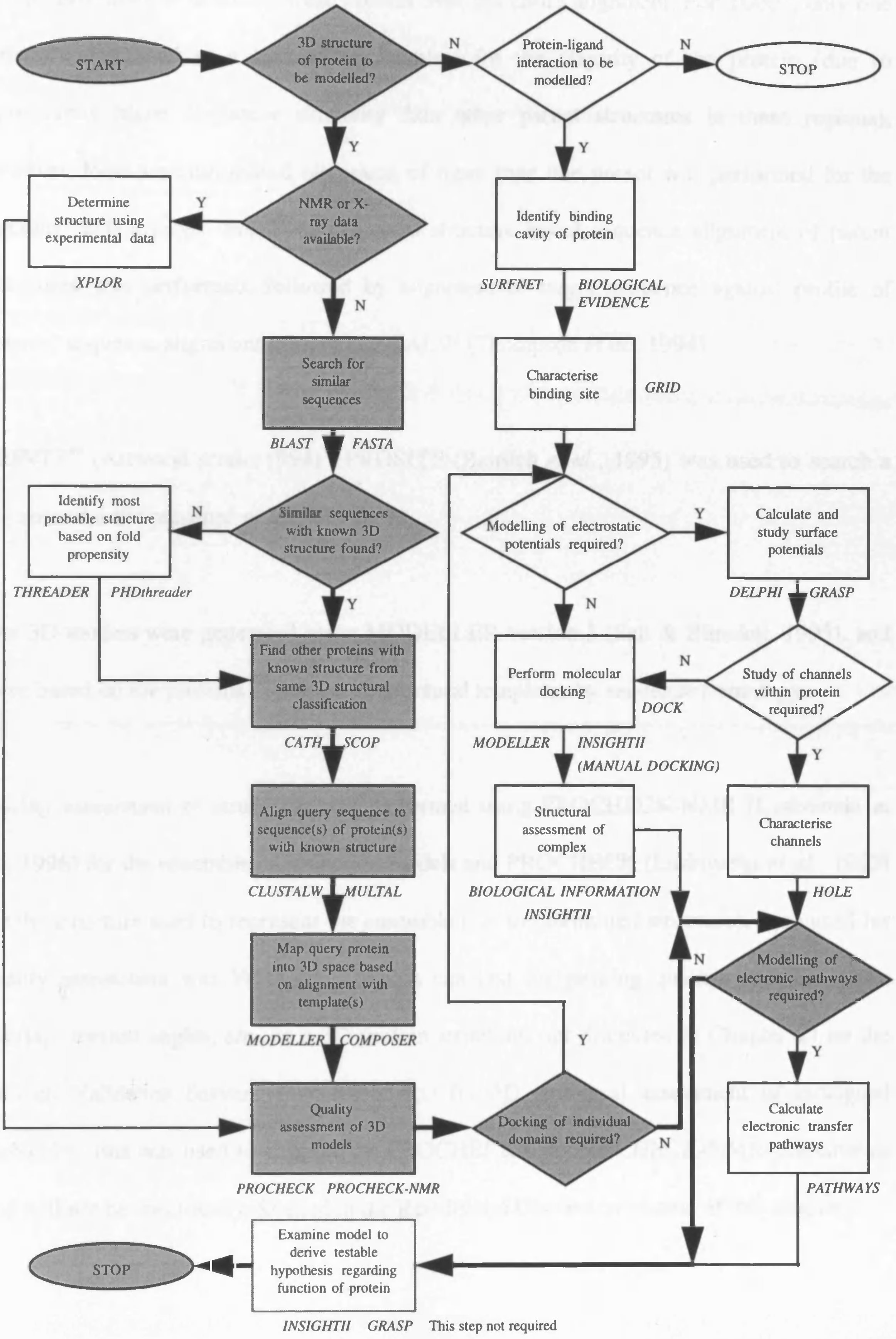

START
3D structure of protein to be modelled?
N
Protein-ligand interaction to be modelled?
N
STOP
Y
Y
Determine structure using experimental data
Y
NMR or X-ray data available?
Identify binding cavity of protein
XPLOR
SURFNET
BIOLOGICAL EVIDENCE
N
Search for similar sequences
Characterise binding site
GRID
BLAST
FASTA
Identify most probable structure based on fold propensity
N
Similar sequences with known 3D structure found?
Modelling of electrostatic potentials required?
Y
Calculate and study surface potentials
THREADER
PHDthreader
Y
N
DELPHI
GRASP
Find other proteins with known structure from same 3D structural classification
Perform molecular docking
N
Study of channels within protein required?
DOCK
CATH
SCOP
MODELLER
INSIGHTII
(MANUAL DOCKING)
Y
Align query sequence to sequence(s) of protein(s) with known structure
Structural assessment of complex
Characterise channels
CLUSTALW
MULTAL
BIOLOGICAL INFORMATION
INSIGHTII
HOLE
Map query protein into 3D space based on alignment with template(s)
N
Modelling of electronic pathways required?
MODELLER
COMPOSER
Y
Y
Quality assessment of 3D models
Docking of individual domains required?
N
Calculate electronic transfer pathways
PROCHECK
PROCHECK-NMR
PATHWAYS
STOP
Examine model to derive testable hypothesis regarding function of protein
INSIGHTII
GRASP
This step not required

Figure 4.1 Flowchart for CASP2 Comparative Modelling (shaded boxes)

and no insertions or deletions were present over the entire alignment. For T0007, only one structure was used as a 'parent' or template for the majority of the protein (due to significantly higher sequence similarity than other parent structures in these regions); however, local structure-based alignment of more than one parent was performed for the calcium binding loops. For T0024, manual structure-based sequence alignment of parent structures was performed, followed by alignment of target sequence against profile of parents' sequence alignment using CLUSTALW (Thompson *et al.*, 1994).

PRINTS[30] (Attwood *et al.*, 1994) / PROSITE (Bairoch *et al.*, 1995) was used to search a sequence for the presence of known motifs.

The 3D models were generated using MODELLER version 3 (Sali & Blundell, 1993), and were based on the proteins identified as structural templates by sequence homology.

Quality assessment of structures was performed using PROCHECK-NMR (Laskowski *et al.*, 1996) for the ensemble of generated models and PROCHECK (Laskowski *et al.*, 1993) for the structure used to represent the ensemble (*i.e.* the submitted structure). Also used for quality assessment was WHATIF[31] (which can test for packing quality, van der Waals overlap, torsion angles, *etc.*, in a 3D protein structure; not discussed in Chapter 2) on the Biotech Validation Server[32], a collaboration for 3D structural assessment of biological molecules; this was used to support the PROCHECK and PROCHECK-NMR assessments and will not be specifically detailed in the Results and Discussion section of this chapter).

[30] URL http://www.biochem.ucl.ac.uk/bsm/dbbrowser/PRINTS/PRINTS.html

[31] URL http://swift.embl-heidelberg.de/whatif/

[32] URL http://biotech.embl-ebi.ac.uk:8400/

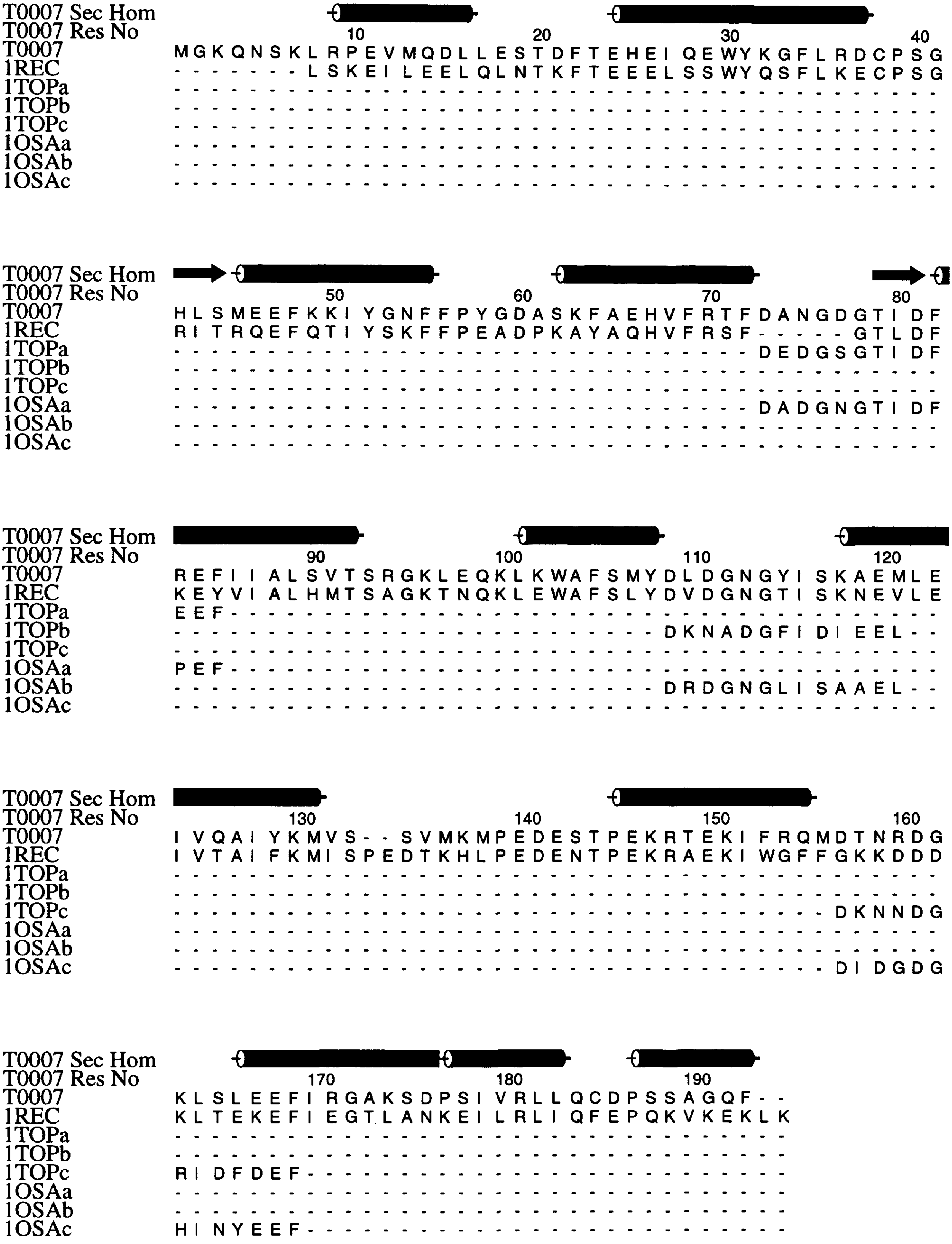
For the purpose of model submission, the selection of a representative structure from the MODELLER-derived ensemble of structures was performed using NMRCLUST (Kelley *et al.*, 1996). NMRCLUST (not discussed in Chapter 2) clusters an ensemble of 3D protein structures with respect to their pairwise RMS distances. Pairwise RMS distances may vary significantly between ensembles of structures (*e.g.* less than 1Å to more than 10Å in NMR-derived structures), and therefore the cutoff distance that defines the clustering threshold is defined automatically by the program. This clustering threshold is dependent on a trade-off between the following factors within the clustering calculation: (1) a 'spread' of points within each particular cluster that is as small as possible, and (2) a total number of clusters which are as few as possible. The structure closest to the centroid of the largest cluster is taken to be the structure that is representative of the whole ensemble.

Positional uncertainties, namely RMSD (root mean square deviation/difference) value across a generated ensemble of models, were calculated for each atom, as required for model submission. This was performed using XPLOR (Brunger, 1996) to generate an average unminimised structure, and then superposing each of the models onto this average structure.

4.3 Results and Discussion

4.3.1 Bovine Neurocalcin Delta (CASP2 Code T0007)

The sequence alignment used for the modelling is shown in Figure 4.2.



```
T0007 Sec Hom
T0007 Res No            10        20        30        40
T0007           MGKQNSKLRPEVMQDLLESTDFTEHEIQEWYKGFLRDCPSG
1REC            -------LSKEILEELQLNTKFTEEELSSWYQSFLKECPSG
1TOPa           -----------------------------------------
1TOPb           -----------------------------------------
1TOPc           -----------------------------------------
1OSAa           -----------------------------------------
1OSAb           -----------------------------------------
1OSAc           -----------------------------------------

T0007 Sec Hom
T0007 Res No           50        60        70        80
T0007           HLSMEEFKKIYGNFFPYGDASKFAEHVFRTFDANGDGTIDF
1REC            RITRQEFQTIYSKFFPEADPKAYAQHVFRSF-----GTLDF
1TOPa           -------------------------------DEDGSGTIDF
1TOPb           -----------------------------------------
1TOPc           -----------------------------------------
1OSAa           -------------------------------DADGNGTIDF
1OSAb           -----------------------------------------
1OSAc           -----------------------------------------

T0007 Sec Hom
T0007 Res No          90       100       110       120
T0007           REFIIALSVTSRGKLEQKLKWAFSMYDLDGNGYISKAEMLE
1REC            KEYVIALHMTSAGKTNQKLEWAFSLYDVDGNGTISKNEVLE
1TOPa           EEF--------------------------------------
1TOPb           --------------------------DKNADGFIDIEEL--
1TOPc           -----------------------------------------
1OSAa           PEF--------------------------------------
1OSAb           --------------------------DRDGNGLISAAEL--
1OSAc           -----------------------------------------

T0007 Sec Hom
T0007 Res No        130         140       150       160
T0007           IVQAIYKMVS--SVMKMPEDESTPEKRTEKIFRQMDTNRDG
1REC            IVTAIFKMISPEDTKHLPEDENTPEKRAEKIWGFFGKKDDD
1TOPa           -----------------------------------------
1TOPb           -----------------------------------------
1TOPc           -----------------------------------DKNNDG
1OSAa           -----------------------------------------
1OSAb           -----------------------------------------
1OSAc           -----------------------------------DIDGDG

T0007 Sec Hom
T0007 Res No       170       180       190
T0007           KLSLEEFIRGAKSDPSIVRLLQCDPSSAGQF--
1REC            KLTEKEFIEGTLANKEILRLIQFEPQKVKEKLK
1TOPa           ---------------------------------
1TOPb           ---------------------------------
1TOPc           RIDFDEF--------------------------
1OSAa           ---------------------------------
1OSAb           ---------------------------------
1OSAc           HINYEEF--------------------------
```

Figure 4.2 Sequence alignment for comparative modelling of neurocalcin delta (CASP2 code T0007), showing secondary structure of homology model. The crystal structure coordinates are not yet available. A single calcium binding region from each of the 1TOP and 1OSA template structures is used for each predicted calcium binding region in neurocalcin delta. Identifiers 'a', 'b' and 'c' are used to denote this.

The protein with highest sequence homology with bovine neurocalcin delta (NCD; CASP2 code T0007) as identified by BLAST was bovine recoverin (calcium sensor in vision; PDB code 1REC, Flaherty *et al.*, 1993) with *p(N)* of 1.2e-68 and percent identity of 67%. The next highest sequence homologies, but with significantly larger values for *p(N)*, were chicken troponin C (skeletal muscle; *p(N)* 2.4e-12; PDB code 1TOP, Satyshur *et al.*, 1994) and *Paramecium tetraurelia* calmodulin (calcium binding protein; PDB code 1OSA, Ban *et al.*, 1994) with *p(N)* of 1.5e-9. Each of these proteins contains 'EF-hand' units (an EF-hand comprises of two α-helices connected by a calcium binding loop).

The locations of the calcium binding regions within NCD were predicted based on the consideration of three approaches: (1) motif search in target and parent proteins using the PRINTS program (occurrences of the EF-hand motif detected *i.e.* oxygen-containing sidechains at residue positions 1, 3, 5, 9 and 12, and a conserved glycine at position 6[33]), (2) extraction of data from Swissprot entries for target and parent proteins, and (3) examination of PDB structures of parent proteins. This information is presented in Table 4.1. Of the four possible regions for calcium binding in these proteins, only three are predicted to do so in NCD based on the results of the motif search. The table also shows that for each of the three NCD calcium binding sites there is at least one parent crystal structure with a bound calcium ion in the corresponding location.

Recoverin (PDB code 1REC) was used as the principal parent (*i.e.* principal structural template for query protein) due to its significantly higher sequence identity with NCD (67%), compared with those of the other two parent proteins (percent identity of these with

[33] The IQ motif (*i.e.* I[/F]QxxxRGxxxRxxΦ, where the first residue in the motif is either isoleucine or phenylalanine, and where Φ is a 'bulky' hydrophobic residue) is found in some calmodulin binding proteins, and is also an EF-hand identifier.

	NCD	1REC	1TOP	1OSA
Calcium binding region 1				
Sequence	R36-F48 RDCPSGHLSMEEF	K37-F49 KECPSGRITRQEF	D30-L42 DADGGGDISTKEL	D20-L32 DKDGDGTITTKEL
EF-hand motif identified? (PRINTS/PROSITE)	No	No	Yes	Yes
Ca^{2+} binding? (Swissprot Data)	Ancestral Ca^{2+} site	Ancestral Ca^{2+} site	Ca^{2+} binding	Ca^{2+} binding
Ca^{2+} ion in PDB file?	N/A	No	No	Yes
Calcium binding region 2				
Sequence	D73-F85 DANGDGTIDFREF	D74-Y86[34] DANSDGTLDFKEY	D66-F78 DEDGSGTIDFEEF	D56-F68 DADGNGTIDFPEF
EF-hand motif identified? (PRINTS/PROSITE)	Yes	Yes	Yes	Yes
Ca^{2+} binding? (Swissprot Data)	Potential Ca^{2+} site	Low Affinity Ca^{2+} binding	Ca^{2+} binding	Ca^{2+} binding
Ca^{2+} ion in PDB file?	N/A	No	No	Yes
Calcium binding region 3				
Sequence	D109-M121 DLDGNGYISKAEM	D110-V122 DVDGNGTISKNEV	D106-L118 DKNADGFIDIEEL	D93-L105 DRDGNGLISAAEL
EF-hand motif identified? (PRINTS/PROSITE)	Yes	Yes	Yes	Yes
Ca^{2+} binding? (Swissprot Data)	Potential Ca^{2+} site	High Affinity Ca^{2+} binding	Ca^{2+} binding	Ca^{2+} binding
Ca^{2+} ion in PDB file?	N/A	Yes	Yes	Yes
Calcium binding region 4				
Sequence	D157-F169 DTNRDGKLSLEEF	K161-F172 KKDDDKLTEKEF	D142-F154 DKNNDGRIDFDEF	D129-F141 DIDGDGHINYEEF
EF-hand motif identified? (PRINTS/PROSITE)	Yes	No	Yes	Yes
Ca^{2+} binding? (Swissprot Data)	Potential Ca^{2+} site	Ancestral Ca^{2+} site	Ca^{2+} binding	Ca^{2+} binding
Ca^{2+} ion in PDB file?	N/A	No	Yes	Yes

Table 4.1 Prediction of calcium binding regions within NCD.

[34] D74-D78 coordinates not present in 1TOP PDB file.

NCD ~30%). However, the calcium binding loops in troponin C (PDB code 1TOP) and calmodulin (PDB code 1OSA) were used as additional template sub-structures for modelling the three predicted calcium binding loops in NCD (the ancestral calcium binding loop in NCD was modelled on recoverin alone). This was performed using calcium binding loop endpoint geometries as references for superposition.

The 3D structures of each calcium binding loop in the parent proteins shows that the calcium ion is hydrogen bonded to five protein atoms and one crystal water molecule. The calcium ions and water molecules comprised part of the template structures. However, these interactions were used to derive distance constraints in order to generate a more accurate model for NCD, since the crystal structure of the principle parent contains bound calcium in only one of the three loops that correspond to a calcium binding loop in NCD. The lower limit for the calcium-water distance constraint was assigned as 1.5Å, since such a calcium-water distance occurs in the template 1OSA. Table 4.2 details the user-specified distance constraints for the NCD model.

Any calcium-Cα distance that is less than 6.5Å in a parent structure is also restrained to its value in order to aid in the positioning of the calcium ion in the NCD model. Also, beta-beta distance constraints (specified in Table 4.2) had to be introduced to maintain the integrity of hydrogen bonds between two anti-parallel β-strands in spatially adjacent EF-hands within the NCD model.

Residues S94-A95, P135-D137 and G160-F172 were removed from the 3D structure of the 1REC template in order to remove consistent structural violations in NCD at the model generation stage.

Calcium ion	Calcium ion constrained to	Upper limit (Å)	Lower limit (Å)
Ca1	D73:OD1, N75:OD1, D77:OD1, T79:O, E84:OE1	3.3	2.0
Ca1	WAT197:OH2	3.3	1.5
Ca2	D109:OD1, D111:OD1, N113:OD1, Y115:O, E120:OE1	3.3	2.0
Ca2	WAT198:OH2	3.3	1.5
Ca3	D157:OD1, N159:OD1, D161:OD1, K163:O, E168:OE1	3.3	2.0
Ca3	WAT199:OH2	3.3	1.5

β-strand atom	β-strand atom	Upper limit (Å)	Lower limit (Å)
I116:N	L164:O	3.3	2.5
I116:O	L164:N	3.3	2.5

Table 4.2 Distance restraints used in model generation of NCD.

An ensemble of 20 models was generated by MODELLER. 13 of these were selected based on lack of significant stereochemical violations. A representative structure was determined from these 13 using NMRCLUST. Figure 4.3 shows the superposition of the representative model and the crystal structure of the principal template structure, 1REC.

Using PROCHECK-NMR across the 20 members of the ensemble, 93% of ϕ/ψ angles were determined to be in 'core' regions of Ramachandran space. The overall G-factor for the representative structure was calculated by PROCHECK to be 0.17 (according to PROCHECK, G-factor scores should ideally be above -0.5 and values below -1.0 may need investigation).

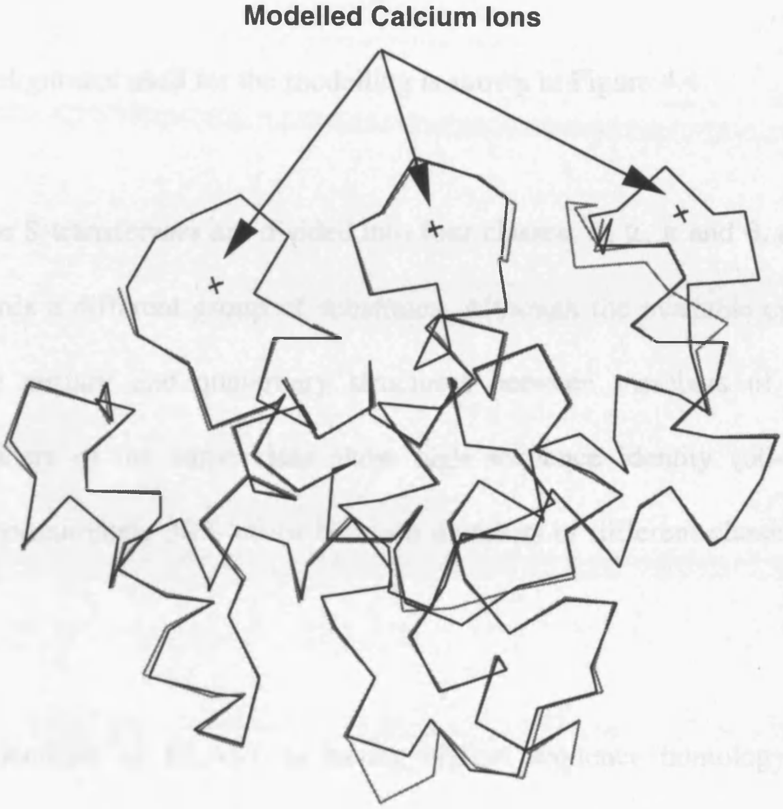


Figure 4.3 Cα superposition of homology model (dark) and crystal structure (light) of Neurocalcin delta (CASP2 code T0007).

At the time of writing, the crystal structure for Neurocalcin Delta (CASP2 code T0007) was not yet solved. Therefore, this protein was not included in the CASP2 analysis, and a comparison cannot be made between the final model and the crystal structure.

4.3.2 Glutathione-S-Transferase (CASP2 Code T0017)

The sequence alignment used for the modelling is shown in Figure 4.4.

The glutathione S-transferases are divided into four classes: α, μ, π and θ, each class being selective towards a different group of substrates. Although the available crystal structures show that the tertiary and quaternary structures between members of all classes are similar[35], members of the same class show high sequence identity (60-90%), whereas identities of approximately 30% occur between members of different classes (Raghunathan *et al.*, 1994).

The protein identified by BLAST as having highest sequence homology with rat liver glutathione transferase (CASP2 code T0017) was rat liver glutathione transferase complexed with a glutathione derivative (PDB code 2GST, Ji *et al.*, 1994) with *p(N)* of 1.5e-142. The protein identified with the next highest sequence homology was the W214F mutant of human muscle glutathione transferase complexed with a glutathione derivative (PDB code 1HNA, Raghunathan *et al.*, 1994) with *p(N)* of 5.7e-129. These two BLAST-identified proteins were rated as significantly more homologous than other proteins. All three of these proteins are from the 'μ' class of glutathione transferases (other BLAST-identified proteins included glutathione-S-transferases from both α and π classes, but these

[35] Crystal structure not currently determined for any class θ glutathione-S-transferase.

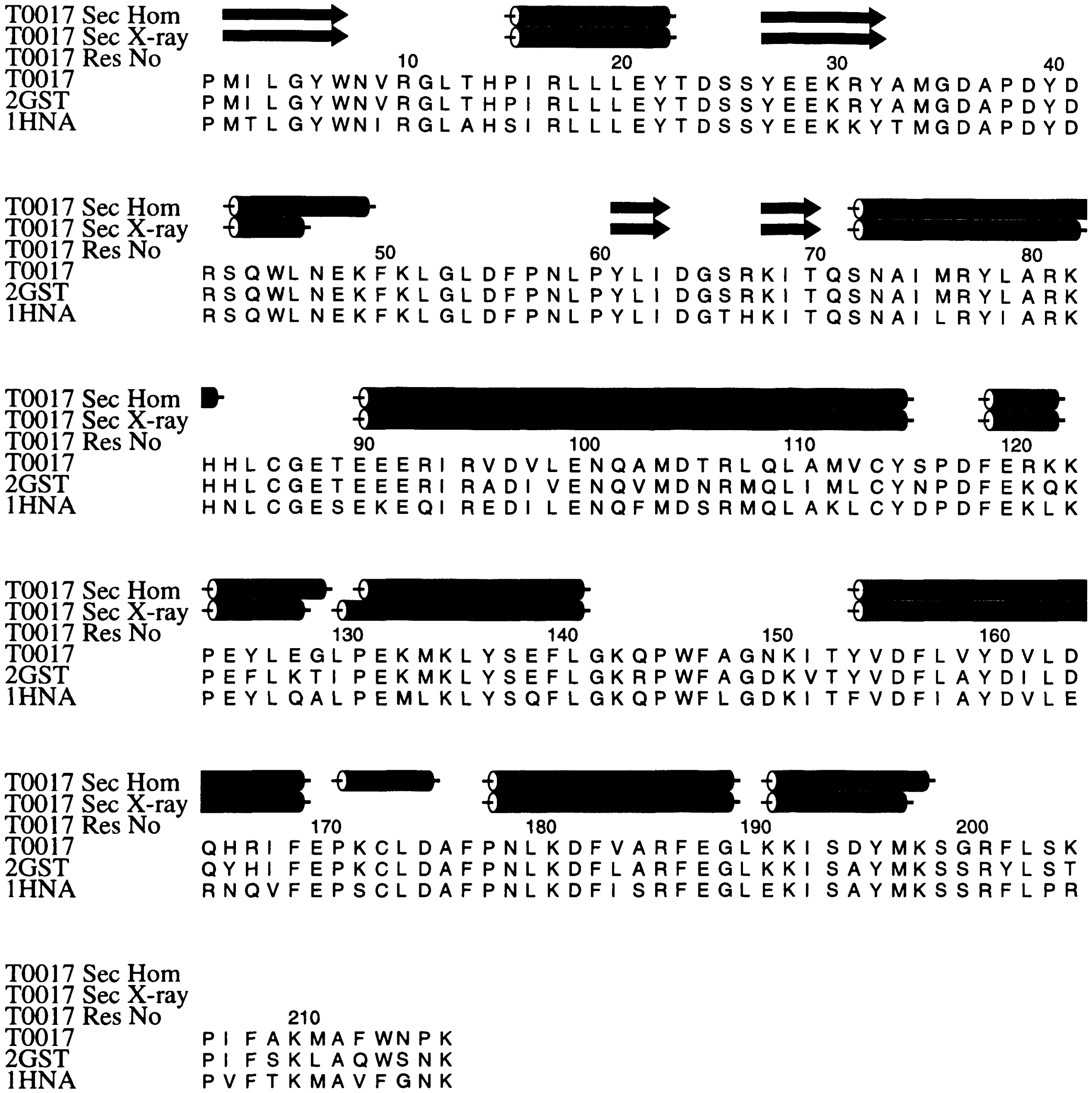

T0017 Sec Hom
T0017 Sec X-ray
T0017 Res No
10 20 30 40
T0017 PMILGYWNVRGLTHPIRLLLEYTDSSYEEKRYAMGDAPDYD
2GST PMILGYWNVRGLTHPIRLLLEYTDSSYEEKRYAMGDAPDYD
1HNA PMTLGYWNIRGLAHSIRLLLEYTDSSYEEKKYTMGDAPDYD
T0017 Sec Hom
T0017 Sec X-ray
T0017 Res No
50 60 70 80
T0017 RSQWLNEKFKLGLDFPNLPYLIDGSRKITQSNAIMRYLARK
2GST RSQWLNEKFKLGLDFPNLPYLIDGSRKITQSNAIMRYLARK
1HNA RSQWLNEKFKLGLDFPNLPYLIDGTHKITQSNAILRYIARK
T0017 Sec Hom
T0017 Sec X-ray
T0017 Res No
90 100 110 120
T0017 HHLCGETEEERIRVDVLENQAMDTRLQLAMVCYSPDFERKK
2GST HHLCGETEEERIRADIVENQVMDNRMQLIMLCYNPDFEKQK
1HNA HNLCGESEKEQIREDILENQFMDSRMQLAKLCYDPDFEKLK
T0017 Sec Hom
T0017 Sec X-ray
T0017 Res No
130 140 150 160
T0017 PEYLEGLPEKMKLYSEFLGKQPWFAGNKITYVDFLVYDVLD
2GST PEFLKTIPEKMKLYSEFLGKRPWFAGDKVTYVDFLAYDILD
1HNA PEYLQALPEMLKLYSQFLGKQPWFLGDKITFVDFIAYDVLE
T0017 Sec Hom
T0017 Sec X-ray
T0017 Res No
170 180 190 200
T0017 QHRIFEPKCLDAFPNLKDFVARFEGLKKISDYMKSGRFLSK
2GST QYHIFEPKCLDAFPNLKDFLARFEGLKKISAYMKSSRYLST
1HNA RNQVFEPSCLDAFPNLKDFISRFEGLEKISAYMKSSRFLPR
T0017 Sec Hom
T0017 Sec X-ray
T0017 Res No
210
T0017 PIFAKMAFWNPK
2GST PIFSKLAQWSNK
1HNA PVFTKMAVFGNK

Figure 4.4 Sequence alignment for comparative modelling of glutathione-S-transferase (CASP2 code T0017), showing secondary structure of homology model (row 1) and crystal structure (row2).

exhibited significantly lower homology with the query sequence than the μ class glutathione-S-transferases).

The 2GST and 1HNA glutathione-S-transferases were therefore selected as the parent proteins, having 85% and 75% identity with the query protein, respectively. No manual refinement of the automatic multiple sequence alignment was required since the homology is very high and no insertions or deletions were present within the alignment.

All glutathione-S-transferases with 'known' 3D structures exist as homodimers. Therefore, T0017 was modelled as such. Although one of the parent structures, 2GST, is stored as a dimer in the PDB databank, the other, 1HNA, is stored as a monomer. The dimer form of 1HNA was generated using the InsightII package based on symmetry information within the 1HNA PDB file.

An ensemble of 20 models was generated by MODELLER (no heteroatoms were modelled). Residues P204 and R205 were deleted from the 1HNA template in order to remove consistent violations in this region in the generated models. No proline exists at residue position 204 in either T0017 or 2GST. However, all three proteins possess a proline residue at position 206 (although the stereochemistry is not conserved, see below).

A representative structure from the ensemble of models was determined using NMRCLUST. Figure 4.5 shows the superposition of the representative model and the crystal structure of GST.

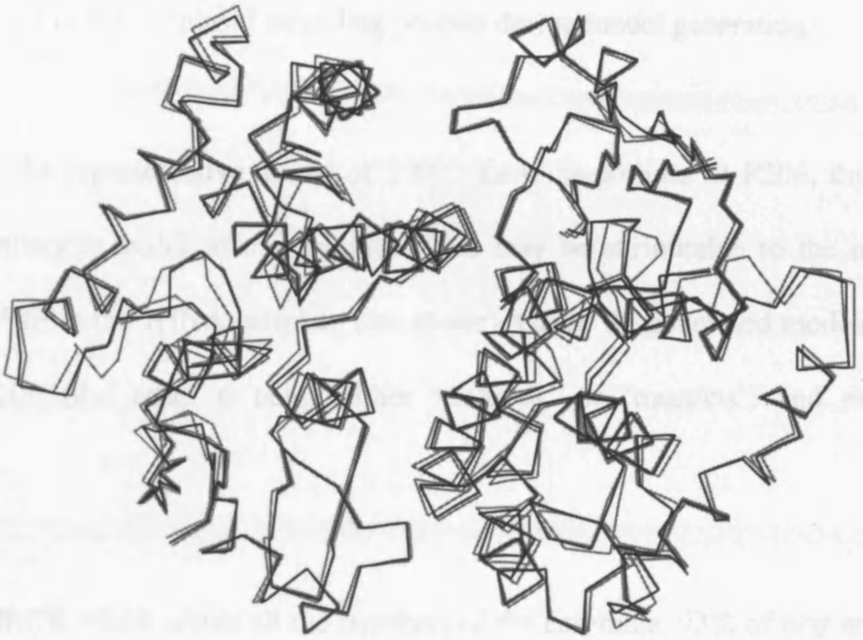


Figure 4.5 Cα superposition of homology model (dark) and crystal structure (light) of the glutathione-S-transferase dimer (CASP2 code T0017).

The 2GST parent structure has cis-prolines at P38, P60 and P206 (both chains), and the 1HNA parent structure has a cis-proline at P60; thus, the stereochemistry differs at the P38 and P206 positions. The most representative model of T0017 has a trans-proline at P38 chain A, and a cis-proline at P38 chain B. Furthermore, of the 20 generated models, 16 were either 'trans-P38 chain A / cis-P38 chain B' or 'cis-P38 chain A / trans-P38 chain B' (only 4 were either 'cis/cis' or 'trans/trans'). This phenomenon is a by-product of the sampling of space in the simulated annealing process during model generation.

Both chains in the representative model of T0017 have cis-prolines at P206, thus echoing the stereochemistry in 2GST at that position. This may be attributable to the removal of P204 and R205 from the 1HNA template (see above). Of the 20 generated models, 14 were 'cis/cis' at P206 (the other 6 being either 'cis/trans' or 'trans/cis', and none being 'trans/trans').

Using PROCHECK-NMR across all the members of the ensemble, 93% of ϕ/ψ angles were determined to be in 'core' regions of Ramachandran space. The overall G-factor for the representative structure was calculated by PROCHECK to be 0.14 (according to PROCHECK G-factor scores should ideally be above -0.5 and values below -1.0 may need investigation).

Table 4.3 shows the structural assessment of this model using some of the CASP2 assessment parameters (Martin *et al.*, 1997). It conveys that the submitted model performed reasonably well for most of the assessment criteria. The positional RMSDs and the percentage of atoms within defined distances between target and model were somewhat better than average. The RMSD for ϕ and ψ mainchain dihedral angles and the χ_1, χ_2 and χ_3

Assessment Parameter	Value for submitted model (Sutcliffe group)	Mean value (standard deviation in parentheses) for submitted models from all participating research groups	Range of values for submitted models from all participating research groups
RMS difference for Cα atoms	0.50 Å	0.89 (0.75) Å	0.45 - 2.73 Å
RMS difference for mainchain and Cβ atoms	0.57 Å	0.95 (0.76) Å	0.49 - 2.77 Å
RMS difference for sidechain atoms	1.53 Å	1.84 (0.99) Å	1.18 - 4.29 Å
RMS difference for all atoms	1.16 Å	1.48 (0.87) Å	0.93 - 3.61 Å
Dihedral angle RMSD for ϕ and ψ angles	17.8°	19.4 (12.2)°	8.6 - 42.2°
Dihedral angle RMSD for χ_1 angles	55.9°	50.0 (12.7)°	38.1 - 75.3°
Dihedral angle RMSD for χ_2 angles	63.0°	56.8 (11.2)°	46.1 - 78.6°
Dihedral angle RMSD for χ_3 angles	74.1°	71.5 (10.3)°	60.3 - 90.4°
'% correct' ϕ and ψ angles in prediction	93 %	92 (8) %	74 - 100 %
'% correct' χ_1 angles in prediction	73 %	74 (13) %	47 - 86 %
'% correct' χ_2 angles in prediction	61 %	66 (13) %	37 - 76 %
'% correct' χ_3 angles in prediction	48 %	46 (12) %	23 - 56 %
% of Cα atoms where ltarget-modell < 1Å	97 %	83 (26) %	23 - 98 %
% of Cα atoms where ltarget-modell < 3Å	100 %	97 (7) %	77 - 100 %
% of all atoms where ltarget-modell < 1Å	82 %	72 (25) %	17 - 89 %
% of all atoms where ltarget-modell < 3Å	97 %	93 (10) %	65 - 98 %
Relative error estimates for Cα atoms	0.37	0.86 (0.24)	0.37 - 1.00
Relative error estimates for mainchain and Cβ atoms	0.37	0.86 (0.24)	0.37 - 1.00
Relative error estimates for sidechain atoms	0.31	0.83 (0.25)	0.31 - 1.00
Relative error estimates for all atoms	0.34	0.84 (0.24)	0.34 - 1.00

A dihedral angle in a model is defined by Martin *et al.* (1997) as 'correct' if its value is equal to that within the crystal structure ±30°.

A relative error estimate is a measure of mean deviation between the observed and estimated distances over all atoms between the target and the model. This measure approaches 0 for correctly estimated errors and 1 for wrong error judgements.

Table 4.3 CASP2 structural assessment of T0017.[36]

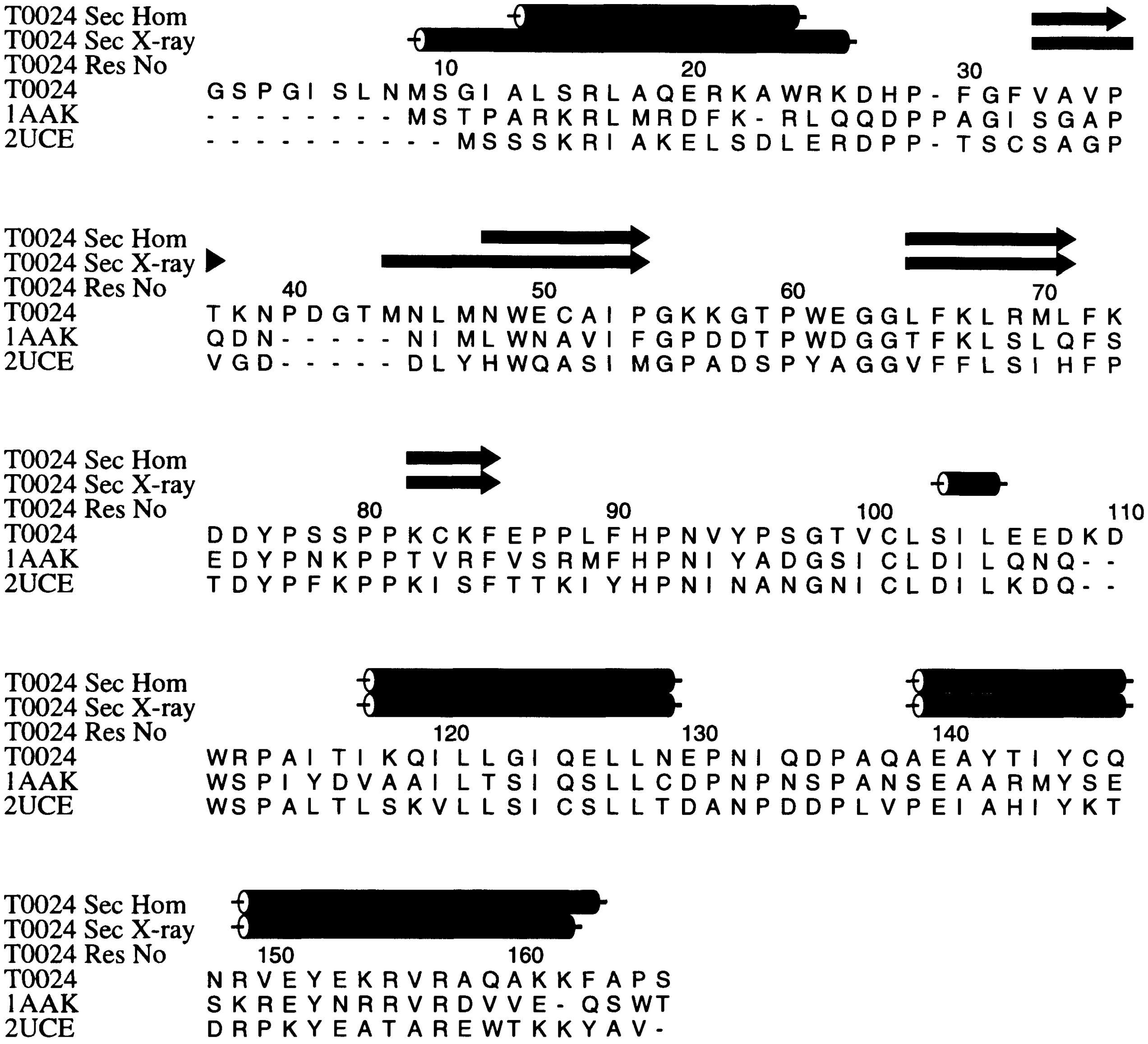
[36] From URL http://www.biochem.ucl.ac.uk/casp2/

sidechain dihedrals were approximately average with respect to angular RMSD and '% correct' dihedral angles. The error estimates were the most accurate compared with the other participating groups.

4.3.3 Ubiquitin Conjugating Enzyme 9 (CASP2 Code T0024)

The protein with highest sequence homology with mouse/human (identical) Ubiquitin Conjugating Enzyme 9 (UBC9; CASP2 code T0024) as identified by BLAST was mouse-ear cress Ubiquitin Conjugating Enzyme 1 (PDB code 1AAK, Cook *et al.*, 1992) with *p(N)* of 4.6e-44. The sequence identity is 37% as determined by the CASP2 assessment (*i.e.* percent identity after a simple pairwise Needleman & Wunsch sequence alignment without manual intervention; see Martin *et al.*, 1997). The parent protein with the next highest sequence homology to the target was yeast Ubiquitin Conjugating Enzyme 4 (PDB code 2UCE, Cook *et al.*, 1993) with *p(N)* of 2.2e-33. There were no other significant BLAST-identified proteins homologous to UBC9, and no other proteins were found to exist with this fold.

The sequence alignment of the query and parent proteins is shown in Figure 4.6. This alignment gives pairwise percent identities of 40% (1AAK) and 36% (2UCE) with UBC9. The first 8 residues in UBC9 are not included within the alignment since they were a by-product of the cloning process, and were excluded from the alignment (and therefore the model). Residues Q29-N31 and Q93-Q95 in 1AAK as well as V27-D29 and K91-Q93 in 2UCE were removed from these template structures in order to remove consistent



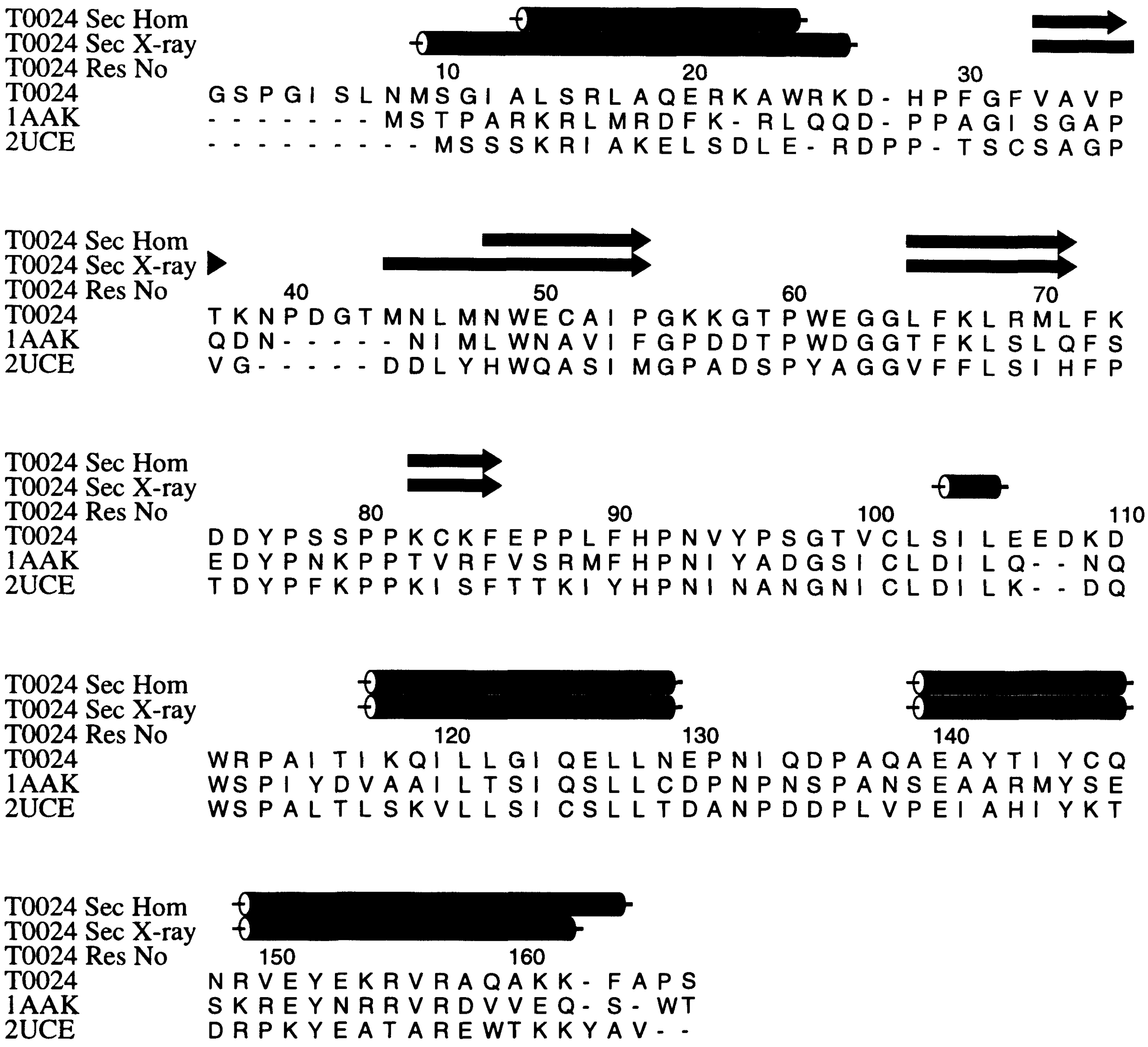
```
T0024 Sec Hom
T0024 Sec X-ray
T0024 Res No              10        20        30
T0024            GSPGISLNMSGIALSRLAQERKAWRKDHP-FGFVAVP
1AAK             --------MSTPARKRLMRDFK-RLQQDPPAGISGAP
2UCE             ----------MSSSKRIAKELSDLERDPP-TSCSAGP

T0024 Sec Hom
T0024 Sec X-ray
T0024 Res No       40        50        60        70
T0024            TKNPDGTMNLMNWECAIPGKKGTPWEGGLFKLRMLFK
1AAK             QDN-----NIMLWNAVIFGPDDTPWDGGTFKLSLQFS
2UCE             VGD-----DLYHWQASIMGPADSPYAGGVFFLSIHFP

T0024 Sec Hom
T0024 Sec X-ray
T0024 Res No          80        90       100       110
T0024            DDYPSSPPKCKFEPPLFHPNVYPSGTVCLSILEEDKD
1AAK             EDYPNKPPTVRFVSRMFHPNIYADGSICLDILQNQ--
2UCE             TDYPFKPPKISFTTKIYHPNINANGNICLDILKDQ--

T0024 Sec Hom
T0024 Sec X-ray
T0024 Res No             120       130       140
T0024            WRPAITIKQILLGIQELLNEPNIQDPAQAEAYTIYCQ
1AAK             WSPIYDVAAILTSIQSLLCDPNPNSPANSEAARMYSE
2UCE             WSPALTLSKVLLSICSLLTDANPDDPLVPEIAHIYKT

T0024 Sec Hom
T0024 Sec X-ray
T0024 Res No      150       160
T0024            NRVEYEKRVRAQAKKFAPS
1AAK             SKREYNRRVRDVVE-QSWT
2UCE             DRPKYEATAREWTKKYAV-
```

Figure 4.6 Submitted sequence alignment for comparative modelling of ubiquitin conjugating enzyme 9 (CASP2 code T0024), showing secondary structure of homology model (row 1) and crystal structure (row 2).

structural violations in the corresponding loop regions in UBC9 at the model generation stage.

The CASP2 assessment showed that the sequence alignment used in modelling the target protein was incorrect. In fact, all research groups performing comparative modelling of this target protein used an incorrect alignment. The Sutcliffe group alignment consisted of 21 incorrectly aligned positions between UBC9 and 1AAK and 19 incorrectly aligned positions between UBC9 and 2UCE; almost all of these involve the N-terminal residues (part of which takes up a helical conformation), the remainder occurring at locations of insertions/deletions, and almost all of the misalignments are out of step by one amino acid position. In correcting the alignment, a little sequence homology is sacrificed, reflecting the over-riding importance of the structure in sequence alignment. The correct alignment is given in Figure 4.7.

An ensemble of 20 models was generated by MODELLER. The two loop regions in UBC9 for which templates residues were omitted (as mentioned above) were not well-packed against the rest of the protein. Also, the Cys-101 sidechain (important in interaction with ubiquitin) in the models was not in an orientation similar to that in the template structures. Therefore, the lowest energy structure from the models where the aforementioned loops were well-packed against the rest of the protein was selected as the sole template for a second round of model generation. Before running the second round of model generation, the Cys-101 sidechain of the new template was manually altered to be in the same conformation as the corresponding cysteine sidechain in 1AAK (the conformation differed slightly between the corresponding cysteines in 1AAK and 2UCE; 1AAK was selected since it possessed higher percentage identity with UBC9). Thus, the second ensemble of 20



```
T0024 Sec Hom
T0024 Sec X-ray
T0024 Res No             10                  20                  30
T0024       G S P G I S L N M S G I A L S R L A Q E R K A W R K D - H P F G F V A V P
1AAK        - - - - - - - M S T P A R K R L M R D F K - R L Q Q D - P P A G I S G A P
2UCE        - - - - - - - - - M S S S K R I A K E L S D L E - R D P P - T S C S A G P

T0024 Sec Hom
T0024 Sec X-ray
T0024 Res No      40                  50                  60                  70
T0024       T K N P D G T M N L M N W E C A I P G K K G T P W E G G L F K L R M L F K
1AAK        Q D N - - - - - N I M L W N A V I F G P D D T P W D G G T F K L S L Q F S
2UCE        V G - - - - - D D L Y H W Q A S I M G P A D S P Y A G G V F F L S I H F P

T0024 Sec Hom
T0024 Sec X-ray
T0024 Res No                  80                  90                 100                 110
T0024       D D Y P S S P P K C K F E P P L F H P N V Y P S G T V C L S I L E E D K D
1AAK        E D Y P N K P P T V R F V S R M F H P N I Y A D G S I C L D I L Q - - N Q
2UCE        T D Y P F K P P K I S F T T K I Y H P N I N A N G N I C L D I L K - - D Q

T0024 Sec Hom
T0024 Sec X-ray
T0024 Res No                       120                 130                 140
T0024       W R P A I T I K Q I L L G I Q E L L N E P N I Q D P A Q A E A Y T I Y C Q
1AAK        W S P I Y D V A A I L T S I Q S L L C D P N P N S P A N S E A A R M Y S E
2UCE        W S P A L T L S K V L L S I C S L L T D A N P D D P L V P E I A H I Y K T

T0024 Sec Hom
T0024 Sec X-ray
T0024 Res No    150                 160
T0024       N R V E Y E K R V R A Q A K K - F A P S
1AAK        S K R E Y N R R V R D V V E Q - S - W T
2UCE        D R P K Y E A T A R E W T K K Y A V - -
```

Figure 4.7 Correct sequence alignment for comparative modelling of ubiquitin conjugating enzyme 9 (CASP2 code T0024), showing secondary structure of homology model (row 1) and crystal structure (row 2).

models was generated by MODELLER based on the manually altered model of UBC9 selected from the first ensemble of models.

A representative structure of the final set of models (*i.e.* second round of generated models) was determined from the ensemble using NMRCLUST. Figure 4.8 shows the superposition of the representative model and the crystal structure of UBC9.

Using PROCHECK-NMR across the members of the final ensemble, 91% of ϕ/ψ angles were determined to be in 'core' regions of Ramachandran space. The overall G-factor for the representative structure was calculated by PROCHECK to be -0.04 (according to PROCHECK G-factor scores should ideally be above -0.5 and values below -1.0 may need investigation).

Table 4.4 shows the structural assessment of this model using some of the CASP2 assessment parameters (Martin *et al.*, 1997). It conveys that the model performed to a similar level with the other participating groups for most of the assessment criteria. The positional RMSDs, the percentage of atoms within defined distances between target and model, RMSD for ϕ and ψ dihedral angles and the '% correct' ϕ and ψ dihedral angles were slightly better than average. For the sidechains, the χ angles in general are not consistently better or worse than average; for example, the χ_1 predictions are worse than average whereas the χ_2 predictions are better than average. The error estimates were slightly more accurate than average.

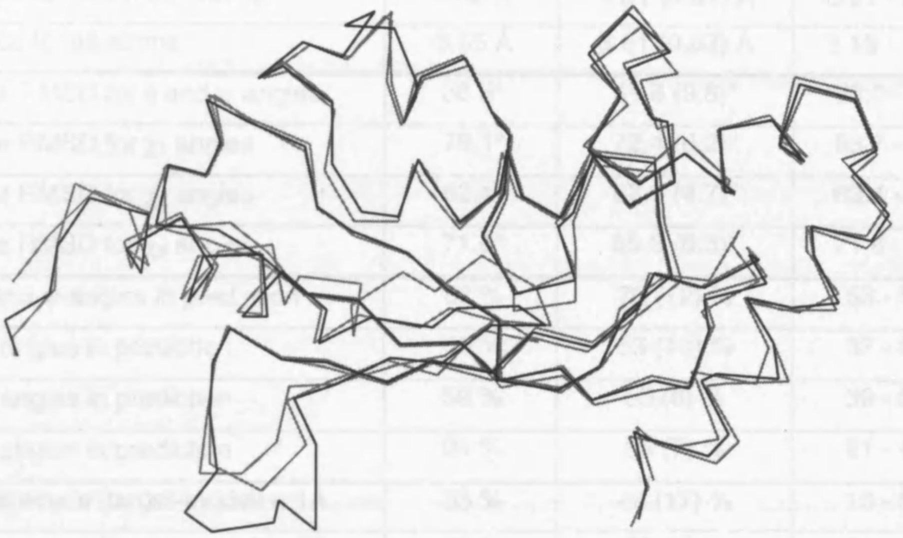


Figure 4.8 Cα superposition of homology model (dark) and crystal structure (light) of ubiquitin conjugating enzyme 9 (CASP2 code T0024).

Assessment Parameter	Value for submitted model (Sutcliffe group)	Mean value (standard deviation in parentheses) for submitted models from all participating research groups	Range of values for submitted models from all participating research groups
RMS difference for Cα atoms	2.80 Å	3.06 (0.68) Å	2.39 - 4.06 Å
RMS difference for mainchain and Cβ atoms	2.79 Å	3.07 (0.69) Å	2.41 - 4.06 Å
RMS difference for sidechain atoms	4.42 Å	4.51 (0.61) Å	3.81 - 5.33 Å
RMS difference for all atoms	3.65 Å	3.81 (0.63) Å	3.15 - 4.60 Å
Dihedral angle RMSD for ϕ and ψ angles	38.5°	41.6 (9.8)°	33.0 - 60.6°
Dihedral angle RMSD for χ_1 angles	79.1°	72.4 (6.2)°	63.7 - 79.1°
Dihedral angle RMSD for χ_2 angles	62.4°	67.9 (4.7)°	62.4 - 75.5°
Dihedral angle RMSD for χ_3 angles	71.6°	85.3 (8.3)°	71.6 - 93.5°
'% correct' ϕ and ψ angles in prediction	83 %	78 (12) %	53 - 86 %
'% correct' χ_1 angles in prediction	50 %	53 (10) %	37 - 65 %
'% correct' χ_2 angles in prediction	59 %	53 (8) %	39 - 59 %
'% correct' χ_3 angles in prediction	34 %	35 (7) %	21 - 41 %
% of Cα atoms where $\lvert$target-model$\rvert$ < 1Å	55 %	46 (17) %	13 - 55 %
% of Cα atoms where $\lvert$target-model$\rvert$ < 3Å	79 %	78 (7) %	64 - 84 %
% of all atoms where $\lvert$target-model$\rvert$ < 1Å	43 %	37 (14) %	10 - 45 %
% of all atoms where $\lvert$target-model$\rvert$ < 3Å	74 %	72 (8) %	55 - 77 %
Relative error estimates for Cα atoms	0.56	0.60 (0.33)	0.27 - 1.00
Relative error estimates for mainchain and Cβ atoms	0.55	0.60 (0.33)	0.26 - 1.00
Relative error estimates for sidechain atoms	0.39	0.55 (0.28)	0.35 - 1.00
Relative error estimates for all atoms	0.48	0.58 (0.30)	0.30 - 1.00

A dihedral angle in a model is defined by Martin *et al.* (1997) as 'correct' if its value is equal to that within the crystal structure ±30°.

A relative error estimate is a measure of mean deviation between the observed and estimated distances over all atoms between the target and the model. This measure approaches 0 for correctly estimated errors and 1 for wrong error judgements.

Table 4.4 CASP2 structural assessment of T0024.[37]

[37] From URL http://www.biochem.ucl.ac.uk/casp2/

4.3.4 CASP2 Assessment of Comparative Modelling

The CASP2 assessment findings presented in this section are extracted from the analysis in Martin *et al.* (1997).

Since the crystal structures are not yet solved for 3 of the 12 target proteins, only 62 of the 81 models (89 of the 122 coordinate sets) could be assessed.

The modelling methods used for CASP2 ranged from fragment-based modelling to approaches based on probability distributions (*e.g.* MODELLER). Almost all methods modelled sidechains by inheritance of χ angles from the parent(s) where possible and, otherwise, by consideration of rotamer libraries, checking for steric overlap and sidechain rotation using various potential functions.

The models in the comparative modelling category generally exhibited relatively similar structural accuracy for a particular target protein. However, direct comparison of results from different groups is not straightforward since not all groups modelled all residues of a given target protein.

Three of the target proteins yielded two independent sets of coordinates, either because the structure was solved in two different crystal forms (*i.e.* T0024 and T0027) or because there are two independent molecules in the assymetric unit of the crystal (*i.e.* T0001). This provided an RMSD value for defining crystal structure accuracy: comparing the two sets gave Cα RMSDs of approximately 0.6Å and all-atom RMSDs of approximately 1.0Å. Most of the models for T0017 were within the Cα crystal structure accuracy, but only half were within the all-atom crystal structure accuracy. None of the T0024 models were within crystal structure accuracy with respect to either Cα or all-atom RMSD (the crystal form with spacegroup $P2_1$ rather than that with spacegroup I222 was used for the assessment since the resolution of the former was slightly higher).

The CASP2 assessment showed that refinement by energy minimisation or molecular dynamics made small improvements to local geometry, and rectified steric clashes and poor chemical interactions, yet generally made insignificant improvements to the RMSD between the model and the crystal structure of the target protein.

The CASP2 assessment found that there was little correlation between the overall geometric quality of a model and the RMSD between the model and the crystal structure of the target protein. Programs for model generation, such as MODELLER, tend to assign values for geometric properties of models based on emprically-derived data. Quality assessment programs, such as PROCHECK, also tend to be based on empirical data. It is therefore important to recognise the difference between precision and accuracy when producing models. However, from Tables 4.3 and 4.4, it is apparent that the errors in the Sutcliffe group's models were estimated fairly well. From the results of the CASP2 assessment, it is anticipated that the more the structure of a model deviates from the parent structure, the less accurate it is likely to be with respect to the target structure. Therefore, it is important to maximise the structural inheritance from the parent to the model. This is simpler where the target and parent sequences have high sequence identity (*e.g.* T0017, sequence identity 85%) than for those with lower sequence identity (*e.g.* T0024, sequence identity 37%).

The most important factor in model accuracy is alignment accuracy. Below a sequence identity of 25%, the alignments were so poor (Cα RMSDs even greater than 18Å) that the models were not considered useful with respect to structural details, even though the overall fold may be accurate. In such cases, the alignment accuracy was found to be as low as 7%. Errors in the sequence alignment used for modelling T0024 were propagated through the whole modelling process for this structure.

The sidechain χ_1 RMSDs are strongly correlated with the Cα RMSDs. For models with Cα deviations less than 1Å, on average 78.5% of sidechains are placed in the 'correct'[38] χ_1 rotamer. However, the χ_1 dihedral angle RMSD for such models is still relatively high. For example, Table 4.3 shows that, for the 'best' T0017 model from the Sutcliffe group (which has an RMSD of 0.50Å between the model and the crystal structure), 73% of χ_1 angles are in the 'correct' rotamer, yet the χ_1 RMSD is 55.9°. For models with Cα RMSDs larger than 2Å, the χ_1 RMSDs are >70° (*i.e.* approaching random). For example, Table 4.4 shows that, for the T0024 model from the Sutcliffe group (which has an RMSD of 2.80Å between the model and the crystal structure), 50% of χ_1 angles are in the 'correct' rotamer, and the χ_1 RMSD is 79.1°.

For the core regions[39], some models have higher RMS distances than the principal parent structure and contain 'serious' structural errors. These errors mainly reflect poor alignments.

[38] A dihedral angle in a model is defined by Martin *et al.* (1997) as 'correct' if its value is equal to that within the crystal structure ±30°.

[39] Core regions are defined as residues where the distance between equivalent target and parent Cα positions is less than 3Å after an iterative structural superposition (see Martin *et al.*, In Press).

For the structurally variable regions (SVRs)[40] the global error is much larger than the local error. This discrepancy arises from two sources: (1) hinge-bending of the loop which will always result in a major difference between global and local RMSDs in loops, and (2) errors in the alignment which often lead to loops being built of the wrong length or shifted in alignment by one or more residues, resulting in a large difference between global and local RMSD or causing both values to be high. The difference between global and local RMSD was greater than 3Å in more than 50% of non-core regions.

The CASP2 assessors attempted to examine whether using multiple parents improved success in modelling SVRs. There were too many variables to consider. However, by examining modelling results by eye, they found that multiple parents could be useful when the homology between target and parent sequences was low, yet did not improve results when sequence homology was high.

4.4 Conclusion

As evident in the Sutcliffe group's models (and in those submitted by other groups), the most important factor in model accuracy is alignment accuracy. For high sequence identity between target and parent sequences (*e.g.* ~85%) where there are no insertions or deletions, it is possible to produce models of crystal structure accuracy. However, lower sequence identity leads to poorer models due to inaccurate sequence alignment.

[40] Structurally variable regions are defined as non-core regions of 3 or more consecutive residues.

Associated with sequence alignment is structural inheritance from parent protein to query protein. Correct structural inheritance from the parent protein is highly important in generating more accurate models. Homology modelling methods which have not solved this problem adequately succeed less well than those that do.

Also, the difference between the local and global RMS distances in loops is significant, and may be attributed to alignment errors and/or hinge-bending within the loops.

Although refinement by energy minimisation or molecular dynamics slightly improve structural quality of models, the RMSD between the model and the crystal structure of the target protein (*i.e.* model accuracy) is not significantly improved.

The assessment of the models submitted by the Sutcliffe group illustrates that the homology modelling methods used within the group and elsewhere in this thesis produce models as good as those produced by the other groups. This gives confidence in the Sutcliffe group's approach to modelling.

CHAPTER 5

NMR SOLUTION STRUCTURE FOR CDC42Hs

5.1 Introduction

Cdc42Hs is a G-protein (guanine nucleotide binding protein or GTPase enzyme) and, as such, acts as a 'molecular switch' for the reaction cascade with which it is associated, cycling between the active (GTP-bound) and inactive (GDP-bound) switch state. Activation occurs through nucleotide exchange of GDP for GTP, whereas deactivation occurs through hydrolysis of GTP. On binding of GTP, specific structural rearrangement in the G-protein is thought to occur which allows it to interact with 'downstream' effectors.

G-proteins occur naturally in either 'large' or 'small' forms. The 'large', or high molecular weight G-proteins, consist of an α, β and γ sub-unit (with the nucleotide binding to the α sub-unit), whereas the 'small', or low molecular weight G-proteins, such as Cdc42Hs, consist of a single subunit.

G-proteins consist of a structurally conserved 'core' into which insertions can be built at particular locations. Such insertions may contribute to structural details that facilitate function-specific interactions.

The *ras* superfamily of G-proteins consists of 6 subfamilies, namely *ras* (including Hras, Kras and Nras), *rho* (including RhoA, RhoB, Rac1, Rac2 and Cdc42Hs), *rab*, *sar*, *ARF* and *ran/TC4*.

Proteins from the *rho* sub-family are involved in various signalling pathways, including those leading to cell mobility, DNA synthesis and cell growth and division. A number of proteins regulate the *rho* subfamily (de-)activation cycle. The rate of *rho* G-protein activation is enhanced through interaction with 'upstream' signalling molecules known as guanine nucleotide exchange factors (GEFs). The signal from the active *rho* G-protein is eventually turned off by intrinsic hydrolysis of the GTP to GDP. This inactivation is enhanced by interaction of the G-protein with GTPase-activating proteins (GAPs). Also enhancing the presence of the inactive form of the *rho* G-protein are molecules known as GDP dissociation inhibitors (GDIs) which repress the dissociation of GDP which is a pre-requisite for nucleotide exchange.

Cdc42Hs is a member of the *rho* sub-family of Ras-like G-proteins, the *rho* subfamily possessing a 13-residue insertion with respect to the *ras* subfamily. Ras is a major regulator of cell growth, and point mutations within this G-protein are found in approximately 30% of human tumours (Scheffzek *et al.*, 1997). Cdc42Hs interacts with a number of effectors, such as the p21-activated serine/threonine kinases (PAKs), Wiskott-Aldrich syndrome protein (WASP), phospholipase D and kinectin, each participating in a particular set of cellular functions.

The objective of this study was to determine the 3D structure of *Homo sapien* Cdc42 (Cdc42Hs) in solution using NMR data, in order to better understand the structure-function relationship of this protein. The modelling structure calculations focus specifically on the 3D structure of the GDP-bound form of Cdc42Hs; the GTP-bound form readily undergoes hydrolysis into the favourable GDP-bound form. However, preliminary NMR information was derived for Cdc42Hs bound to a GTP analogue during the calculation of 3D structures for Cdc42Hs·GDP. The study was performed in collaboration with Prof. R.E. Oswald and co-workers (Department of Pharmacology, College of Veterinary Medicine, Cornell University, USA).

Information on *rho* subfamily G-protein structures previous to this work consisted of a homology model of Cdc42Hs (Sutcliffe *et al.*, 1994) based on GDP- and GTP-analogue bound Hras crystal structures (*ras* subfamily).

Three crystal structures have recently been solved for *rho* subfamily G-proteins; firstly, mutant (F78S) Rac1 protein bound to a non-hydrolysable GTP analogue (PDB code 1MH1; Hirshberg *et al.*, 1997), secondly, Cdc42Hs bound to a non-hydrolysable GTP analogue and in complex with p50RhoGAP (PDB code 1AM4, on hold until 22 Jun 1998; Rittinger *et al.*, 1997a), and thirdly, RhoA bound to a GTP analogue and in a transition state complex with RhoAGAP, or RhoA-specific GAP (PDB code 1TX4, on hold until 29 Jul 1998; Rittinger *et al.*, 1997b). Structural comparisons between the latter two crystal structures and the Cdc42Hs NMR-derived ensemble of structures cannot therefore be described.

5.2 Methods

The methodological approach to this study is described by the flowchart in Figure 5.1. The methods are detailed in Chapter 2.

All protein expression, purification and NMR experiments were carried out by Prof. R.E. Oswald & co-workers (Department of Pharmacology, College of Veterinary Medicine, Cornell University, USA). These collaborators assigned and analysed the spectra and supplied distance constraints for structure calculation. Experiments designed to generate dihedral angle restraints were unsuccessful, and therefore no experimentally-derived dihedral angle restraints were used in the model generation. The primary NMR data is listed in the appendix of this thesis.

NOESY-derived distance constraints were classified into five categories according to the relative strength of the NOESY cross-peak: <2.4Å, <2.9Å, <3.4Å, <4.0Å and <5.5Å. Also, two constraints were assigned per hydrogen bond (these were incorporated into the distance constraints list): the amide proton to carbonyl oxygen distance was constrained to between 1.8Å and 2.3Å, and the amide nitrogen to carbonyl oxygen between 2.5Å and 3.3Å. In the absence of experimentally-derived restraints, dihedral angle restraints were assigned for residues involved in α-helices and β-strands; these were 'loosely' constrained (due to the lack of any experimental restraints) to 'favourable' regions of ϕ/ψ space' (*i.e.* α-helix, ϕ: -80 ± 50°, ψ: -20 ± 50°; β-strand, ϕ: -105 ± 65°, ψ: -145 ± 45°) as derived from the PROCHECK data set (Laskowski *et al.*, 1993).

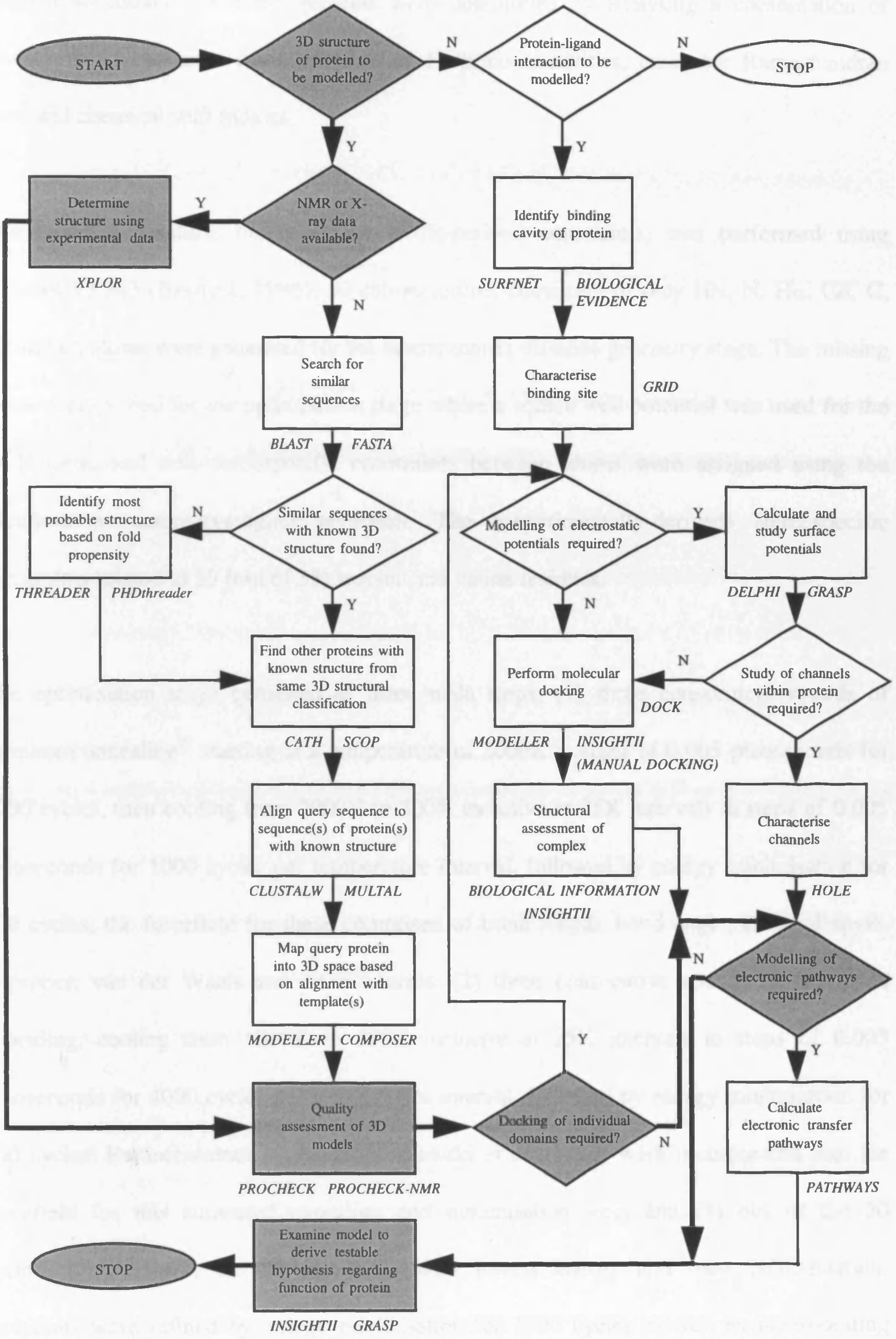

START
3D structure of protein to be modelled?
N
Protein-ligand interaction to be modelled?
N
STOP
Y
Y
Determine structure using experimental data
Y
NMR or X-ray data available?
Identify binding cavity of protein
XPLOR
SURFNET
BIOLOGICAL EVIDENCE
N
Search for similar sequences
Characterise binding site
GRID
BLAST
FASTA
Identify most probable structure based on fold propensity
N
Similar sequences with known 3D structure found?
Modelling of electrostatic potentials required?
Y
Calculate and study surface potentials
THREADER
PHDthreader
Y
N
DELPHI
GRASP
Find other proteins with known structure from same 3D structural classification
Perform molecular docking
N
Study of channels within protein required?
DOCK
CATH
SCOP
MODELLER
INSIGHTII
(MANUAL DOCKING)
Y
Align query sequence to sequence(s) of protein(s) with known structure
Structural assessment of complex
Characterise channels
CLUSTALW
MULTAL
BIOLOGICAL INFORMATION
INSIGHTII
HOLE
Map query protein into 3D space based on alignment with template(s)
N
Modelling of electronic pathways required?
MODELLER
COMPOSER
Y
Y
Quality assessment of 3D models
Docking of individual domains required?
N
Calculate electronic transfer pathways
PROCHECK
PROCHECK-NMR
PATHWAYS
STOP
Examine model to derive testable hypothesis regarding function of protein
INSIGHTII
GRASP

Figure 5.1 Flowchart of Methods for Modelling Cdc42Hs (shaded boxes)

Regular secondary structural elements were determined by analysing a combination of amide proton exchange rates, patterns of NOE connectivities, ensemble Ramachandran plots and chemical shift indices.

The model generation, based on the NMR-derived constraints, was performed using XPLOR v3.843 (Brunger, 1996). 50 sub-structures consisting of only HN, N, Hα, Cα, C, Cβ and Cγ atoms were generated for the metric matrix distance geometry stage. The missing atoms were added for the optimisation stage where a square well potential was used for the NOE term, and non-stereospecific constraints between atoms were assigned using the pseudo-atom centre-averaging approach. The (experimentally-derived) stereospecific constraints related to 29 (out of 35) leucine and valine residues.

The optimisation stage consisted of three main steps, (1) three consecutive rounds of simulated annealing[41] starting at a temperature of 2000K in steps of 0.005 picoseconds for 1000 cycles, then cooling from 2000K to 300K inclusive at 25K intervals in steps of 0.005 picoseconds for 1000 cycles per temperature interval, followed by energy minimisation for 200 cycles; the forcefield for these comprised of bond length, bond angle, dihedral angle, improper, van der Waals and NOE[42] terms, (2) three consecutive rounds of simulated annealing, cooling from 1000K to 300K inclusive at 25K intervals in steps of 0.003 picoseconds for 4000 cycles per temperature interval, followed by energy minimisation for 200 cycles; Ramachandran restraints (Kuszewski *et al.*, 1996) were incorporated into the forcefield for this simulated annealing and minimisation step, and (3) out of the 50 optimised structures, the 20 structures with lowest energy and least NOE-restraint violations were refined by energy minimisation for 1000 cycles as well as incorporating

[41] It was found that three consecutive simulated annealing steps resulted in an improved set of optimised structures while not resorting to an unnecessarily high CPU time.

[42] The NOE term included hydrogen bond constraints.

chemical shift restraints for Cα, Cβ and Hα atoms (Kuszewski *et al.*, 1995a, 1995b) into the forcefield.

Proton-proton distances which violated the NOE distance constraints consistently across the resulting optimised ensemble of structures were examined by the collaborators and the constraints list updated as necessary. The violation of a distance constraint could arise from one of three possibilities, which were addressed in the following order: (1) If the violation corresponded to a misassignment, it was reassigned. (2) If the violation did not correspond to a misassignment but could have arisen as a result of the misassignment of another peak, it was left in the constraint set. (3) In several cases, some ambiguity was present due to the inability to assign all of the sidechains. In these few cases, the constraint was removed. In addition, modelled structures were used to help resolve previously ambiguous NOE assignments. The new constraints list was used to calculate a new set of structures, and this entire procedure was then repeated iteratively in order to overcome consistent structural violations.

The quality of structures was assessed using and PROCHECK-NMR (Laskowski *et al.*, 1996) for the final ensemble of structure and PROCHECK (Laskowski *et al.*, 1993) for a single, representative structure.

NMRCLUST (Kelley *et al.*, 1996; see Methods section of Chapter 3 for brief description) was used to select the representative structure from the final ensemble of 20 models. The well-defined or 'core' atoms across the ensemble were determined using NMRCORE (Kelley *et al.*, 1997; not discussed in Chapter 2). This determination is based on the similarity of dihedral angle values across an ensemble of protein structures. The threshold that distinguishes core atoms from non-core atoms is derived automatically. The program also determines clusters of structurally rigid regions, or LSDs (Local Structural Domains), that exist as a subset of the well-defined atoms, since structurally well-defined regions within proteins may exist as independent rigid structures separated by flexible regions. The threshold that is used to cluster the LSDs is determined automatically, based on pairwise interatomic distances within each structure. It is worth noting that loop regions in proteins, as well as helical and sheet regions, can be well-defined across an ensemble. Therefore, the calculations performed by NMRCORE are indiscriminate and consistent over all regions of protein structure.

5.3 Results and Discussion

Cdc42Hs is a protein consisting of 191 residues. The protein was expressed in *Escherichia coli* as part of a fusion protein with glutathione S-transferase. The cleavage of the fusion protein occurred within the *E. coli* glutathione S-transferase sequence, 7 residues before the N-terminus of Cdc42Hs. The Cdc42Hs was also cleaved 4 residues before the C-terminus during this process. The sample protein therefore consisted of 194 residues, numbered from -7 to 187 (no residue was assigned with number '0').

Approximately 1850 distance constraints were obtained from NOESY spectra. In addition to these distance constraints, 50 hydrogen bonds were included. Approximately 200 dihedral angle restraints were also assigned for residues involved in α-helices and β-strands. The NOEs, CSI, amide exchange rates and secondary structure are summarised in Figure 5.2. The alignment and packing of the β-sheet is illustrated in Figure 5.3.

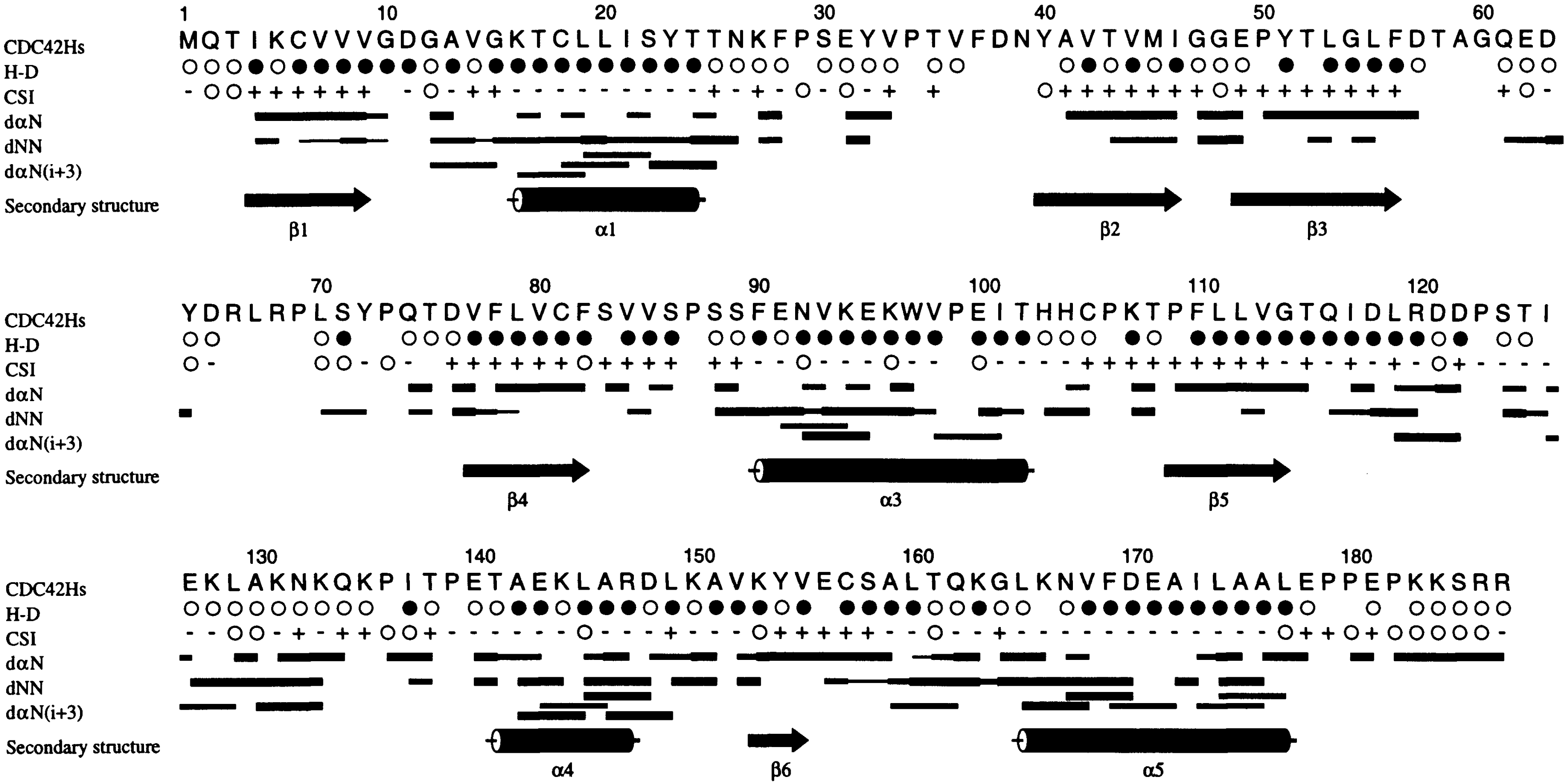

1 10 20 30 40 50 60
CDC42Hs
MQTIKCVVVGDGAVGKTCLLISYTTNKFPSEYVPTVFDNYAVTVMIGGEPYTLGLFDTAGQED
H-D
CSI
dαN
dNN
dαN(i+3)
Secondary structure
β1
α1
β2
β3
70 80 90 100 110 120
YDRLRPLSYPQTDVFLVCFSVVSPSSFENVKEKWVPEITHHCPKTPFLLVGTQIDLRDDPSTI
β4
α3
β5
130 140 150 160 170 180
EKLAKNKQKPITPETAEKLARDLKAVKYVECSALTQKGLKNVFDEAILAALEPPEPKKSRR
α4
β6
α5

Figure 5.2 Summary of assignments and secondary structure of Cdc42Hs. Row 1: sequence of Cdc42Hs. Row 2: amide protons exhibiting deuterium exchange (H-D) in D_2O; half-lives of greater than 1 hour are shown using closed circles, half-lives of less than 1 hour are shown using open circles. Rows 3,4: composite CSIs of the Hα, Cα, Cβ and C' protons in the GDP-bound and GMPPCP-bound forms, where '-' is correlated with helix, '+' is correlated with strand, and an empty circle is correlated with coil. Rows 5,6,7: $d_{\alpha N}$, d_{NN} and $d_{\alpha N(i+3)}$ NOE intensities, where thick bars indicate strong NOEs (<2.9 Angstroms), medium bars indicate intermediate NOEs (<4 Angstroms), and thin bars indicate weak NOEs (<5.5 Angstroms). Row 8: secondary structural elements, specifying helices and strands.

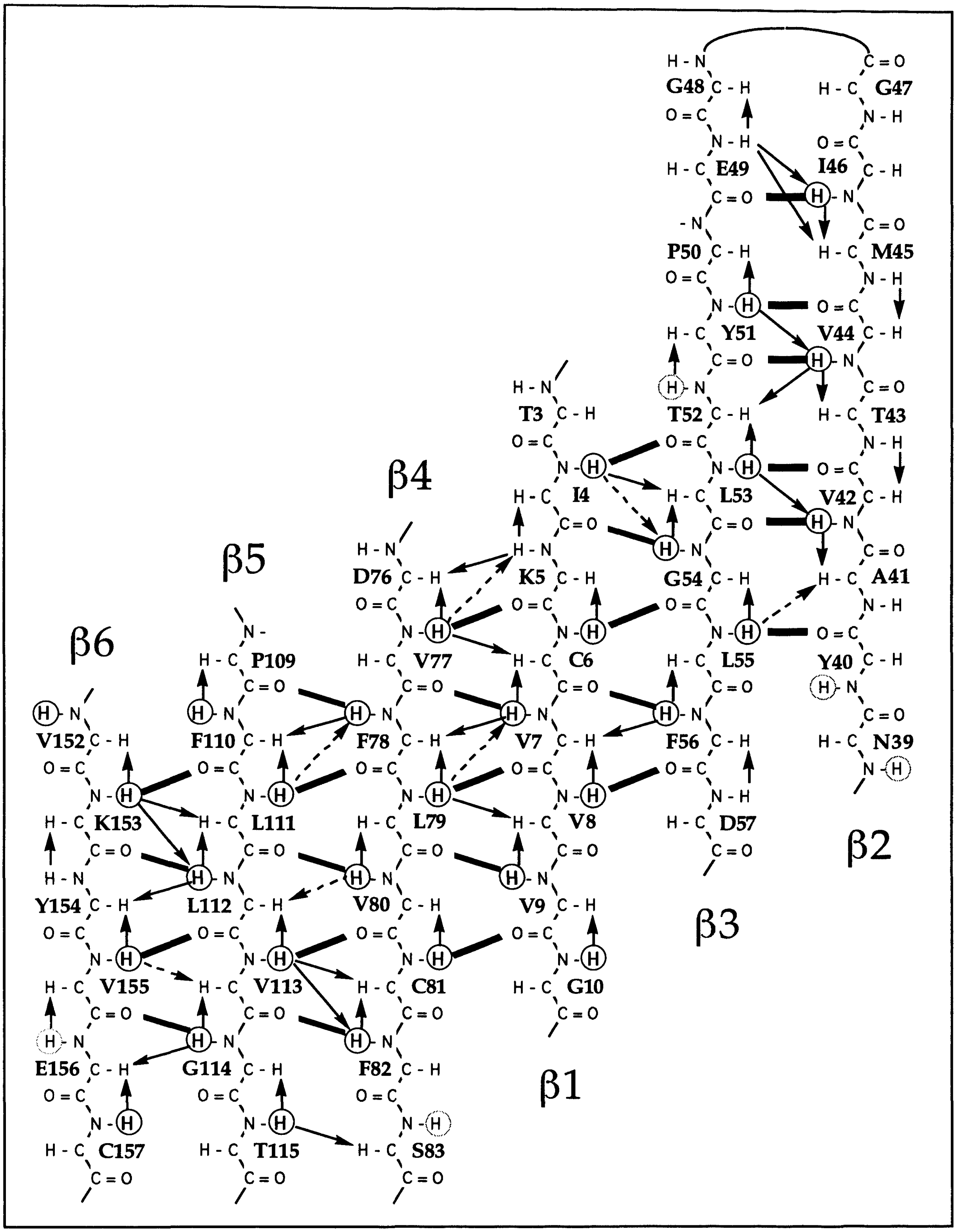


Figure 5.3 Alignment and packing of the β-sheet in Cdc42Hs. H-bonds are indicated by heavy bars; strong NOEs (<2.9Å) are indicated by solid arrows, and weak NOEs (<5.5Å) are indicated by dashed arrows. Amide protons exhibiting deuterium exchange with half-lives greater than 1 hour in D_2O are indicated by dark solid circles, whereas light solid circles indicate amide protons for which it was not possible to measure exchange rates due to peak overlap.

The total number of distance constraints for the Cdc42Hs structure is approximately 2100, that is, approximately 11 constraints per residue. No consistent NOE violations above 1Å were present across the ensemble of modelled structures, and almost all consistent NOE violations (average of 4 violations per structure) above 0.5Å were overcome.

Due to a lack of constraints (only 4 determined) between the GDP and the protein, it was not possible to constrain the GDP within the calculated Cdc42Hs structures. Therefore, the NMR solution structure does not contain GDP. However, the position of the GDP was determined approximately by superposing the appropriate Cα atoms of the Hras•GDP crystal structure (PDB code 1Q21; Tong *et al.*, 1991) onto the 'core' Cα atoms of each of the 20 Cdc42Hs structures using XPLOR, then transposing the position of the GDP from the Hras•GDP crystal structure onto Cdc42Hs (the 'core' residues, as determined by NMRCORE, are residue numbers 3-9, 16-28, 41-47, 49-56, 77-85, 88-102, 109-114, 117, 118, 141-176). XPLOR was then used for the energy minimisation of the Cdc42Hs•GDP complex; the atoms in the protein were initially kept fixed in order to remove steric overlap within the complex. (Hras was the only *ras* superfamily G-protein with known 3D structure at the time of modelling; no structures for *rho* subfamily members were known.)

The NMR-derived ensemble of structures for Cdc42Hs are described in Feltham *et al.*, (1997). The coordinates for the ensemble of structures are publically available from the PDB databank (accession code 1AJE).

5.3.1 Description of Structures

The 3D structure of Cdc42Hs (Figure 5.4) consists of a central six-stranded β-sheet, with packing order β2-β3-β1-β4-β5-β6, where all the strands are parallel except β2. Helices α1 and α5 are on the 'concave', nucleotide binding side of the sheet, and helices α3 and α4 are on its 'convex' side.

The quality of structures for the Cdc42Hs model is described in Table 5.1. The protein 'core' is well-defined across the ensemble, as shown in Figure 5.5a, and has an RMSD of 0.57 ± 0.08Å (excluding hydrogen atoms). The rms deviation across the ensemble over all atoms (excluding hydrogen atoms) is 3.8 ± 0.5Å.

The non-'core' atoms across the ensemble cover five main regions within the protein. These are (1) the N-terminal region which includes the 7 non-native residues, (2) the C-terminal region after the α5 helix, (3) the region corresponding to 'Switch I' in Hras (*i.e.* residues 31-40 in Cdc42Hs), (4) the region corresponding to 'Switch II' in Hras (*i.e.* residues 57-74 in Cdc42Hs), and (5) an insertion sequence with respect to Hras (*i.e.* residues 122-134 in Cdc42Hs).

The switch regions in the Hras crystal structure (PDB code 1Q21; Tong *et al.*, 1991) also show a higher level of disorder than the rest of the structure. This was also observed in the region of the Rac1·GMPPNP crystal structure (PDB code 1MH1; Hirschberg *et al.*, 1997) corresponding to Switch I. While structural disorder in the switch regions of Cdc42Hs are primarily caused by a lack of constraints (since they contain some residues which could not be assigned), ^{15}N relaxation studies (relating to variability in N-H bond directionality)

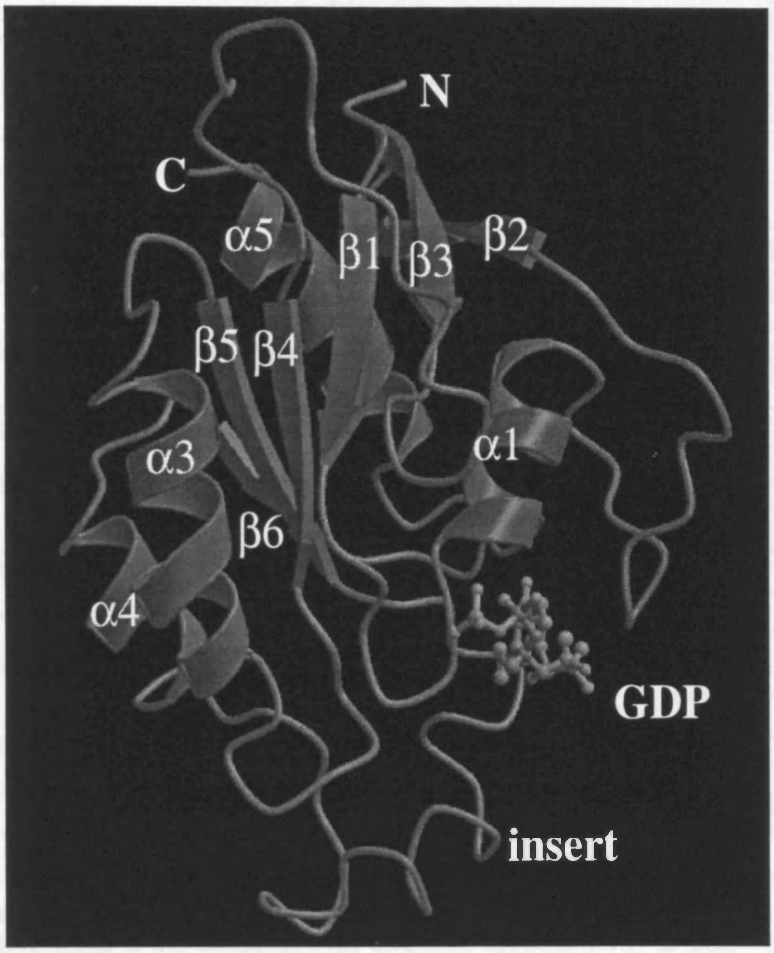


Figure 5.4 Representative structure of GDP-bound Cdc42Hs in ribbon representation with secondary structural elements, the insert region, and the N- and C-termini indicated.

Structure(s) used to represent ensemble	Cdc42Hs$_{clust}$*	Cdc42Hs$_{lowE}$*	Cdc42Hs$_{Av(SD)}$ *
Ramachandran Plot Statistics**			
Residues in:			
most favoured regions	59.9%	61.0%	62.9(2.9)%
additional allowed regions	29.1%	29.7%	28.2(2.7)%
generously allowed regions	8.1%	7.0%	6.2(1.5)%
disallowed regions	2.9%	2.3%	2.7(1.2)%
χ_1 Std. dev. from ideal (degrees)	25.8	24.1	24.7(0.8)
ω angle Std. dev. from ideal (degrees)	1.5	1.3	1.4(0.1)
No. heavy atoms with ≥10% vdW overlap	78	78	80.0(10.6)
No. heavy atoms with ≥20% vdW overlap	2	0	0.7(0.8)
No. of nonbonded contacts < 2.1A	0	0	0.0(0.0)
No. of nonbonded contacts < 2.3A	2	1	1.2(0.8)
No. of nonbonded contacts < 2.5A	12	9	10.8(2.2)
Overall G-factor***	-0.57	-0.58	-0.59(0.02)

* Cdc42Hs$_{clust}$ is the most representative structure (structure 18) as defined by the program NMRCLUST, Cdc42Hs$_{lowE}$ is the lowest energy structure (structure 15) as defined by XPLOR, and Cdc42Hs$_{Av(SD)}$ indicates the average across all 20 structures with standard deviation in parentheses.

** The percentage of residues falling into the four regions of the Ramachandran plot defined by PROCHECK (Laskowski *et al.*, 1993).

*** Ideally, G-factor scores should be above -0.5.

Table 5.1 Stereochemical quality of structures for Cdc42Hs.

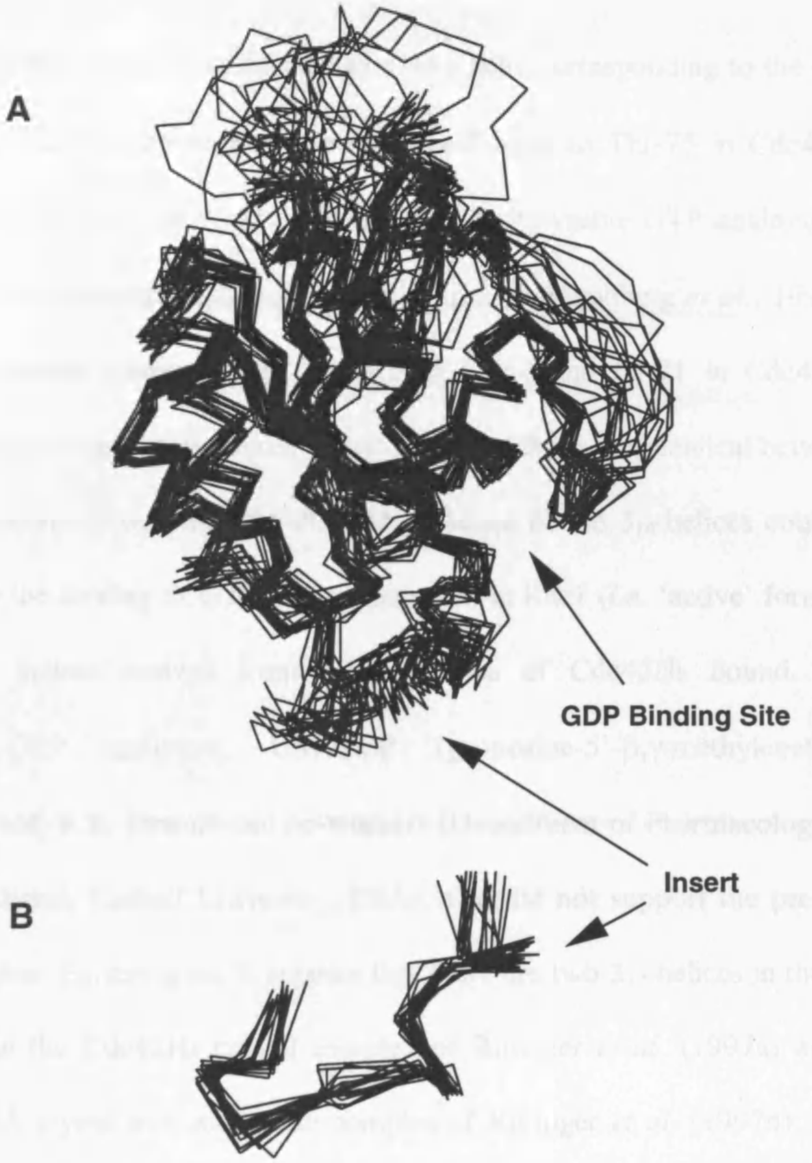
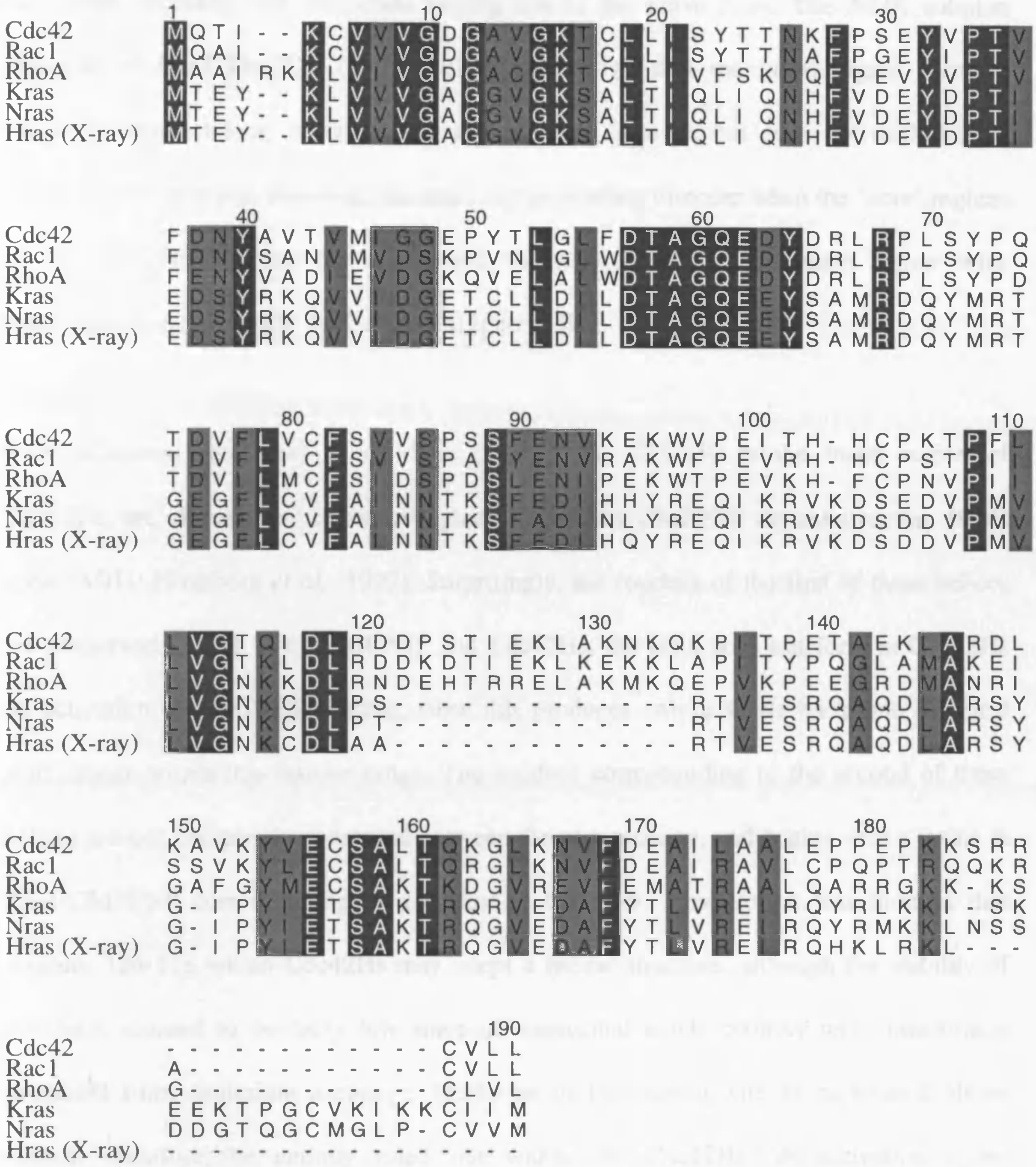


Figure 5.5 Cα superposition of Cdc42Hs ensemble of structures (A), and 'Insert' region residues aligned against each other (B).

currently underway will confirm whether or not these regions in Cdc42Hs are more flexible than the remainder of the protein. This suggests that flexibility in the switch regions may be characteristic of *ras*-like proteins in solution.

The Cdc42Hs NMR data exhibits no evidence for a helix corresponding to the α2 α-helix in Hras (the Hras α2 helix corresponds to residues Pro-69 to Thr-75 in Cdc42Hs). In the crystal structure of Rac1 complexed with the non-hydrolysable GTP analogue, GMPPNP (guanosine-5'-β,γ-imidotriphosphate), PDB code 1MH1 (Hirshberg *et al.*, 1997), there are two short 3_{10}-helices corresponding to residues 62-64 and 68-71 in Cdc42Hs. This is perhaps surprising since the sequences of Cdc42Hs and Rac1 are identical between residues 57 and 79 inclusive. It was thought that the presence of the 3_{10}-helices could have been associated with the binding of GTP (or its analogue) in Rac1 (*i.e.* 'active' form). However, chemical shift indices derived from NMR studies of Cdc42Hs bound to the non-hydrolysable GTP analogue, GMPPCP (guanosine-5'-β,γ-methylenetriphosphate), performed by Prof. R.E. Oswald and co-workers (Department of Pharmacology, College of Veterinary Medicine, Cornell University, USA), also did not support the presence of any helix in this region. Furthermore, it appears that there are two 3_{10}-helices in this vicinity for Cdc42Hs within the Cdc42Hs crystal complex of Rittinger *et al.* (1997a) and for RhoA within the RhoA crystal transition state complex of Rittinger *et al.* (1997b). It is possible that the presence of the 3_{10}-helices in the crystal structures and their absence in the NMR structure of these *rho* subfamily G-proteins may be attributed to conformational difference in this region between crystal and solution states.

A 13-residue insertion exists in Cdc42Hs (residues 122-134) with respect to the Hras sequence (see Figure 5.6), and is present only in the *rho* sub-family of *ras*-like G-proteins.



```
                1                     10                  20                  30
Cdc42           M Q T I - - K C V V V G D G A V G K T C L L I S Y T T N K F P S E Y V P T V
Rac1            M Q A I - - K C V V V G D G A V G K T C L L I S Y T T N A F P G E Y I P T V
RhoA            M A A I R K K L V I V G D G A C G K T C L L I V F S K D Q F P E V Y V P T V
Kras            M T E Y - - K L V V V G A G G V G K S A L T I Q L I Q N H F V D E Y D P T I
Nras            M T E Y - - K L V V V G A G G V G K S A L T I Q L I Q N H F V D E Y D P T I
Hras (X-ray)    M T E Y - - K L V V V G A G G V G K S A L T I Q L I Q N H F V D E Y D P T I

                      40                  50                  60                  70
Cdc42           F D N Y A V T V M I G G E P Y T L G L F D T A G Q E D Y D R L R P L S Y P Q
Rac1            F D N Y S A N V M V D S K P V N L G L W D T A G Q E D Y D R L R P L S Y P Q
RhoA            F E N Y V A D I E V D G K Q V E L A L W D T A G Q E D Y D R L R P L S Y P D
Kras            E D S Y R K Q V V I D G E T C L L D I L D T A G Q E E Y S A M R D Q Y M R T
Nras            E D S Y R K Q V V I D G E T C L L D I L D T A G Q E E Y S A M R D Q Y M R T
Hras (X-ray)    E D S Y R K Q V V L D G E T C L L D I L D T A G Q E E Y S A M R D Q Y M R T

                          80                  90                  100                  110
Cdc42           T D V F L V C F S V V S P S S F E N V K E K W V P E I T H - H C P K T P F L
Rac1            T D V F L I C F S V V S P A S Y E N V R A K W F P E V R H - H C P S T P I I
RhoA            T D V I L M C F S I D S P D S L E N I P E K W T P E V K H - F C P N V P I I
Kras            G E G F L C V F A I N N T K S F E D I H H Y R E Q I K R V K D S E D V P M V
Nras            G E G F L C V F A I N N T K S F A D I N L Y R E Q I K R V K D S D D V P M V
Hras (X-ray)    G E G F L C V F A I N N T K S F E D I H Q Y R E Q I K R V K D S D D V P M V

                                120                   130                 140
Cdc42           L V G T Q I D L R D D P S T I E K L A K N K Q K P I T P E T A E K L A R D L
Rac1            L V G T K L D L R D D K D T I E K L K E K K L A P I T Y P Q G L A M A K E I
RhoA            L V G N K K D L R N D E H T R R E L A K M K Q E P V K P E E G R D M A N R I
Kras            L V G N K C D L P S - - - - - - - - - - - - - R T V E S R Q A Q D L A R S Y
Nras            L V G N K C D L P T - - - - - - - - - - - - - R T V E S R Q A Q D L A R S Y
Hras (X-ray)    L V G N K C D L A A - - - - - - - - - - - - - R T V E S R Q A Q D L A R S Y

                150                 160                 170                 180
Cdc42           K A V K Y V E C S A L T Q K G L K N V F D E A I L A A L E P P E P K K S R R
Rac1            S S V K Y L E C S A L T Q R G L K N V F D E A I R A V L C P Q P T R Q Q K R
RhoA            G A F G Y M E C S A K T K D G V R E V F E M A T R A A L Q A R R G K K - K S
Kras            G - I P Y I E T S A K T R Q R V E D A F Y T L V R E I R Q Y R L K K I - S K
Nras            G - I P Y I E T S A K T R Q G V E D A F Y T L V R E I R Q Y R M K K L N S S
Hras (X-ray)    G - I P Y I E T S A K T R Q G V E D A F Y T L V R E I R Q H K L R K L - - -

                                          190
Cdc42           - - - - - - - - - - - - C V L L
Rac1            A - - - - - - - - - - - C V L L
RhoA            G - - - - - - - - - - - C L V L
Kras            E E K T P G C V K I K K C I I M
Nras            D D G T Q G C M G L P - C V V M
Hras (X-ray)    - - - - - - - - - - - - - - - -
```

Figure 5.6 Sequence alignment of six *ras*-superfamily G-proteins, where identical residues are shown on a black background and similar residues on a grey background. The first three sequences are from the *rho*-subfamily and last three are from the *ras*-subfamily. The GDP from the Hras crystal structure was used as a reference for modelling GDP into Cdc42Hs.

Homology modelling studies of the GDP- and GTP-bound forms of Cdc42Hs based on Hras crystal structures (Sutcliffe *et al.*, 1994) indicated that the insert region might act as a large flap, occluding the nucleotide binding site in the active form. The NMR solution structure of Cdc42Hs·GDP (*i.e.* inactive form) shows that the insert region forms a relatively compact loop structure, and is positioned close to, but does not occlude, the nucleotide binding site. However, the insert region exhibits disorder when the 'core' regions of the Cdc42Hs ensemble are superposed; the superposition of the insert region alone shows that it is structurally well-defined (Figure 5.5b).

Two α-helices, equivalent to residues 117-120 and 123-130 in the insert region of Cdc42Hs, are observed in the insert region of the Rac1·GMPPNP crystal structure (PDB code 1MH1; Hirshberg *et al.*, 1997). Surprisingly, the residues of the first of these helices are conserved in both Rac1·GMPPNP and Cdc42Hs; the helix does not form in Cdc42Hs on activation by GMPPCP binding, since this produces only a single backbone chemical shift change within this residue range. The residues corresponding to the second of these helices are not completely conserved between the two proteins, and begins with a lysine in Rac1·GMPPNP corresponding to a proline in Cdc42Hs. However, it was thought that residues 124-128 within Cdc42Hs may adopt a helical structure, although the stability of this helix seemed to be fairly low since no associated amide protons were measurably protected from deuterium exchange. Occlusion of the binding site as mentioned above cannot, therefore, be entirely ruled out within the Cdc42Hs (de-)activation cycle. Nevertheless, it appears that, in Rittinger *et al.* (1997a), the second of these helices is present in the Cdc42Hs crystal complex, and that, in Rittinger *et al.* (1997b), it is present in the RhoA crystal transition state complex, whereas the first of the helices is present in neither of these.

5.3.2 Effector Interaction and (De-)Activation Cycle Regulation

Indication of conformational changes in the conversion of the inactive to the active form of Cdc42Hs for the purpose of effector binding were suggested by changes in chemical shift between the binding of GDP and the binding of GMPPCP (non-hydrolysable GTP analogue). Chemical shift changes were experienced in both Switch I and Switch II, as well as the loop between β1 and α1 (the P-loop). Another region, designated as 'Switch III', consisting of the β4-α3 loop and the α3 helix also experienced chemical shift changes. The chemical shifts within the insert region of Cdc42Hs are almost entirely invariant, suggesting a lack of conformational change between the inactive and active forms. The chemical shift changes in Cdc42Hs are shown in Figure 5.7, which illustrates a large, contiguous 'switch surface', suggesting conformational change associated with protein-protein interaction.

Mutation studies for the investigation in the interaction between G-proteins and GAPs have yielded evidence for the 'arginine finger hypothesis', where it is proposed that an arginine residue in the GAP stabilises the transition state in its reaction with the G-protein. Mutation of the arginine finger have shown significantly decreased GAP activity in RhoGAP (Unpublished results reported by Rittinger *et al.*, 1997a) and RasGAP (Ahmadian *et al.*, 1997). Structure-based studies of the interaction between *rho* subfamily G-proteins and RhoGAPs has recently been presented (Rittinger *et al.*, 1997a and 1997b), as well as that of between $G_{\alpha i1}$, the α-subunit of a trimeric G-protein complexed with RGS4, an associated regulatory protein (Tesmer *et al.*, 1997), providing structural evidence for this hypothesis. The Cα positions of the arginine finger residues of RhoAGAP (Arg-85) and $G_{\alpha i1}$ (Arg-178) are 10Å apart, but the guanidinium groups in the arginine sidechains superpose to RMSDs of under 1Å (Rittinger *et al.*, 1997b).

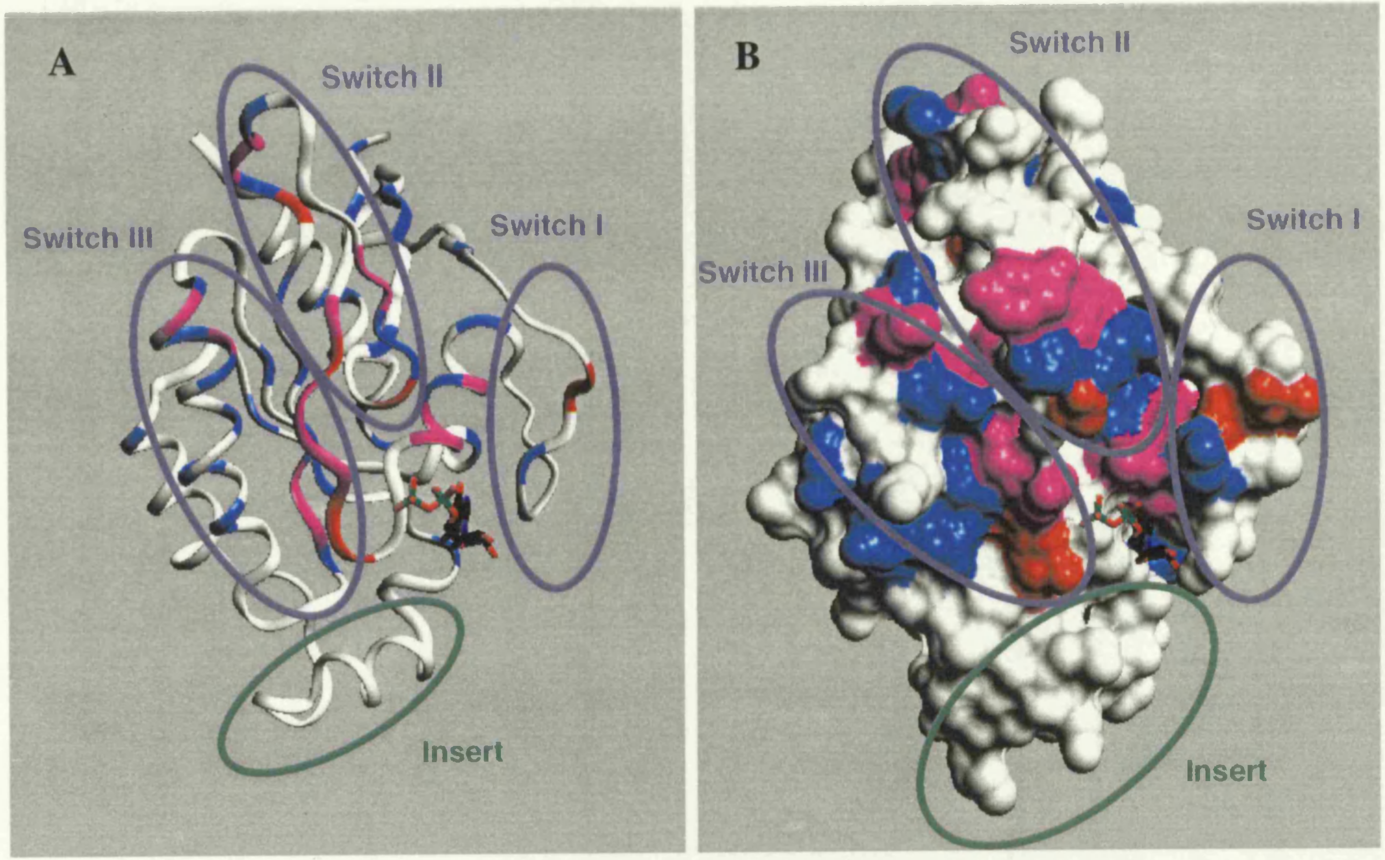


Figure 5.7 The GDP-bound structure of Cdc42Hs using ribbon (A) and space-filled (B) representations to illustrate the locations of the switch regions, the insert region and the chemical shift changes between the GDP-bound and GMPPCP-bound states. In both cases, GDP is shown using a stick representation. The colouring scheme of the protein (both A and B) indicates the degree to which chemical shift changes are observed: large (red), medium (magenta) and small (blue).

In the crystal structure complex Cdc42Hs·GMPPNP/p50rhoGAP (Rittinger *et al.*, 1997a), the arginine finger residue, Arg-85, in p50RhoGAP makes H-bonding interactions with the mainchain of the conserved Gly-12 in the P-loop of Cdc42Hs. However, there is a 20° rotation between this ground state complex Cdc42Hs·GMPPNP/p50rhoGAP of Rittinger *et al.* (1997a) and the RhoA·GDP·AlF_4^-/p50rhoGAP crystal structure transition state complex (Rittinger *et al.*, 1997b)[43] which contributes to transition state stabilisation. In the transition state complex Arg-85 from p50rhoGAP interacts with one of the fluorides of AlF_4^- (equating to a γ-phosphate oxygen) and with the β-γ phosphate bridging oxygen (Rittinger *et al.*, 1997b). In the transition state, Gln-63 in Switch II of RhoA hydrogen bonds to a water molecule that is directly in line with the β-oxygen of GDP and the aluminium ion (Rittinger *et al.*, 1997b).

Although the insert region is not discussed in Rittinger *et al.* (1997a and 1997b), Switches I and II and the P-loop of the *rho* proteins are shown to be important in the interaction with a shallow pocket in p50rhoGAP which is lined with conserved residues within rhoGAPs. The protein-protein interface is comprised of an extensive network of H-bonds, as well as hydrophobic packing interactions. Within the Switch I region of Cdc42Hs, Val-36 and Phe-37 in particular are involved in hydrophobic packing interactions with residues in p50rhoGAP (Rittinger *et al.*, 1997a). In Switch II of Cdc42Hs, Tyr-64 is involved in H-bonds with p50rhoGAP while Leu-67 is involved in hydrophobic packing against p50RhoGAP (Rittinger *et al.*, 1997a). Within the Switch I region of RhoA, the mainchain carbonyl of Tyr-34 interacts with Asn-194 in the p50rhoGAP (Rittinger *et al.*, 1997b). In

[43] The Cdc42Hs·GDP·AlF_4^-/p50rhoGAP transition state complex could not be crystallised. However, the high sequence homology of these *rho* subfamily G-proteins (*i.e.* Cdc42Hs and RhoA) justifies this comparison.

Switch II of RhoA, Asp-65 and Tyr-66 are involved in an extensive H-bonding network with residues in p50rhoGAP (Rittinger *et al.*, 1997b).

The insert region of the Rho subfamily may be involved in the mechanism of RhoGDI action. No known GDIs exist for *ras* subfamily G-proteins, and this could be attributed to the fact that *ras* proteins lack the insert region present in *rho* subfamily G-proteins (Wu *et al.*, 1997). Wu *et al.* (1997) therefore performed mutation studies on the insert region of Cdc42Hs, using a short sequence of Ras in its place. Firstly, studies were performed for mutation effects on Cdc42GAP activity; no significant difference in activity was found between wild type and mutant Cdc42Hs, thus showing that the insert was not essential for Cdc42GAP activity. When testing for GDI activity, it was found that significantly lower activity in RhoGDI occurred with the mutant than with the wild type, conveying GDI sensitivity to changes in the insert region of Cdc42Hs; however, the binding affinity of both mutant and wild type Cdc42Hs to RhoGDI were essentially identical.

Therefore, it is possible that RhoGDI could induce the insert region of Cdc42Hs to occlude the nucleotide binding site (in order to inhibit GDP dissociation, thus enhancing the presence of the inactive form of Cdc42Hs). Occlusion of the nucleotide binding site is an idea first proposed by Sutcliffe *et al.* (1994). Another possibility is that the insert region undergoes conformational change to stabilise the nucleotide through specific binding interactions in order to fulfil the effect of the GDI.

5.4 Conclusion

The 'core' residues across the Cdc42Hs ensemble, as determined by NMRCORE, are residue numbers 3-9, 16-28, 41-47, 49-56, 77-85, 88-102, 109-114, 117, 118, 141-176. The non-'core' atoms across the ensemble cover regions that include Switch I (residues 31-40), Switch II (residues 57-74) and the 13-residue 'insert' (residues 122-134) with respect to the *ras* subfamily of G-proteins. However, the 13-residue insert in Cdc42Hs is compact and well-defined locally.

Chemical shift changes between the active (GTP-bound) and inactive (GDP-bound) conformations of Cdc42Hs, when mapped onto the 3D structure of the GDP-bound form, showed conformational change over a large and contiguous 'switch surface' (comprising of Switch I, Switch II and the newly designated Switch III), but not within the 13-residue insert region. This also negates an earlier homology model-based prediction that the insert region occludes the nucleotide binding site after G-protein activation for the purpose of specificity of interaction with its downstream effector or with its relevant GAP protein which enhances de-activation.

While the results of this NMR-based structural study indicate that the Cdc42Hs insert region does not undergo conformational change for the purpose of interaction with its effector or GAP protein, a recent mutation study on Cdc42Hs raises the idea that the insert region undergoes conformational change on interaction with rhoGDI. This conformational change in Cdc42Hs could be to either (1) occlude its nucleotide binding site, or (2) stabilise the GDP *via* specific binding interactions with the GDI. Both hypotheses rationalise the inhibition of GDP dissociation; however, further study is required to determine the role of the insert region in Cdc42Hs.

CHAPTER 6

CHARACTERISATION OF LIGAND BINDING IN THE THIRD DOMAIN OF THE HUMAN HOMOLOGUE OF DROSOPHILA DISCS-LARGE TUMOUR SUPPRESSOR PROTEIN (PDZ-3_{Dlg})

6.1 Introduction

PDZ domains are named after three proteins in which they are found, that is, Post-synaptic density protein (PSD-95; $M_r = 95K$), Discs large tumour suppressor protein (Dlg) and the epithelial tight junction protein, Zonula occludens-1 (ZO1).

PDZs are associated with junctions between biological cells. They are distinct domains of approximately 90-100 residues in length which can be found in proteins implicated in ion channel and receptor clustering (cellular receptors are not distributed randomly over the cell surfaces but are clustered into regions, *e.g.* receptors at nerve synapses).

Certain PDZ domains (including the one presented in this study) interact specifically with the Thr/Ser-X-Val tri-peptide motif ('X' denotes any amino acid) at the C-terminus of target proteins (*i.e.* receptors, ion channels). However, some PDZ domains are specific to C-termini containing the tetra-peptide motif Glu-Thr/Ser-X-Val/Ile, whereas other PDZs target C-termini with aromatic and hydrophobic residues. Furthermore, a number of PDZ domains appear to mediate homodimerisation, leading to the formation of an intact protein molecule, thus conveying PDZ-PDZ interaction specificity in these cases.

The tumour suppressor gene APC (adenomatous polyposis coli gene) is mutated in the majority of sporadic colorectal tumours and in most cases of familial adenomatous polyposis (FAP), a dominantly inherited disease characterised by multiple adenomatous polyps in the colon. The human homologue of *Drosophila* discs-large tumour suppressor protein (Dlg) is implicated in binding the C-terminal region of the APC gene product (Matsumine *et al.*, 1996). The Dlg-APC complex is thought to participate in the regulation of both cell cycle progression and neuronal function. Dlg consists of three individual PDZ domains, and two other distinct domains; the two non-PDZ domains are an SRC homology 3 (SH3) non-enzymatic signalling domain and a guanylate kinase-like catalytic domain. The protein studied in this work is the third PDZ domain of the human homologue of *Drosophila* discs-large tumour suppressor protein (PDZ-3_{Dlg}).

The objective of this study was to model the tri-peptide Thr-Asp-Val-COO$^-$ (*i.e.* an instance of the C-terminal peptide motif Thr/Ser-X-Val) into the 3D structure of PDZ-3_{Dlg} in order to gain a structural insight, and therefore a functional rationale, into the PDZ-ligand interaction. To this end, Arg-471 in PDZ-3_{Dlg}, corresponding to a predominantly conserved positive residue position in PDZ domains, was used as a reference for carboxylate binding of the ligand, and Gly-475 to Phe-478 in PDZ-3_{Dlg}, corresponding to the largely conserved GLGF motif in PDZ domains, as used as a reference for hydrophobic interactions with the valine sidechain in the ligand.

The study was performed in collaboration with Prof. R.C. Liddington and co-workers (Department of Biochemistry, University of Leicester, UK), who determined the crystal structure of the 96-residue protein, PDZ-3_{Dlg} (PDB code 1PDR; Cabral *et al.*, 1996).

6.2 Methods

The methodological approach to this study is described by the flowchart in Figure 6.1. The methods are detailed in chapter 2.

SURFNET (Laskowski, 1995) was used to identify cavities on the surface of the 3D structure of the PDZ-3_{Dlg} domain. A grid separation of 0.8Å was used.

GRID (Goodford, 1985) was then used in the largest cavity to determine the interaction energies between the PDZ-3_{Dlg} structure and various probes that represented the chemical constituents of the tri-peptide ligand (*i.e.* COO^-, OH, CH_3, C=O and NH). A grid separation of 0.2Å was used. The energy contour maps output from GRID were used as a guide to manually dock the ligand into PDZ-3_{Dlg}. Suitable distance constraints between the PDZ-3_{Dlg} domain and the ligand were selected based on the GRID contour maps.

These constraints then formed the basis of the automatic modelling of the ligand into PDZ-3_{Dlg} using MODELLER (Sali & Blundell, 1993). An ensemble of 10 models of the complex was generated (the MODELLER-derived PDZ-3_{Dlg} model was determined using the crystal structure as a template; the structural parameters for the ligand are already present within the MODELLER library files). The modelled complex with 'lowest energy' was superposed onto the PDZ-3_{Dlg} crystal structure, producing the 'final structure'.

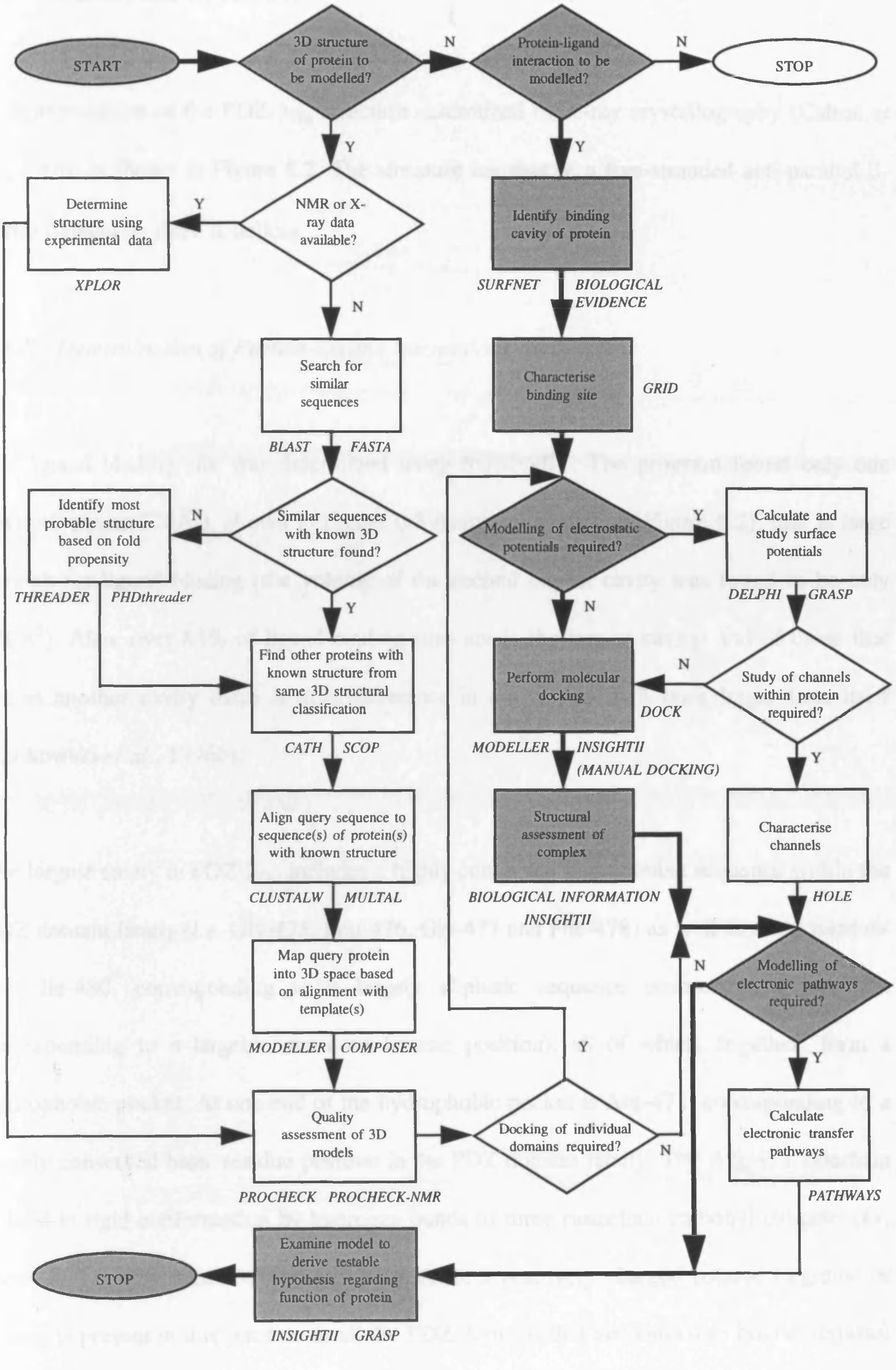

START
3D structure of protein to be modelled?
N
Protein-ligand interaction to be modelled?
N
STOP
Y
Y
Determine structure using experimental data
Y
NMR or X-ray data available?
Identify binding cavity of protein
XPLOR
SURFNET
BIOLOGICAL EVIDENCE
N
Search for similar sequences
Characterise binding site
GRID
BLAST
FASTA
Identify most probable structure based on fold propensity
N
Similar sequences with known 3D structure found?
Modelling of electrostatic potentials required?
Y
Calculate and study surface potentials
THREADER
PHDthreader
Y
N
DELPHI
GRASP
Find other proteins with known structure from same 3D structural classification
Perform molecular docking
N
Study of channels within protein required?
DOCK
CATH
SCOP
MODELLER
INSIGHTII
(MANUAL DOCKING)
Y
Align query sequence to sequence(s) of protein(s) with known structure
Structural assessment of complex
Characterise channels
CLUSTALW
MULTAL
BIOLOGICAL INFORMATION
INSIGHTII
HOLE
Map query protein into 3D space based on alignment with template(s)
N
Modelling of electronic pathways required?
MODELLER
COMPOSER
Y
Y
Quality assessment of 3D models
Docking of individual domains required?
N
Calculate electronic transfer pathways
PROCHECK
PROCHECK-NMR
PATHWAYS
STOP
Examine model to derive testable hypothesis regarding function of protein
INSIGHTII
GRASP

Figure 6.1 Flowchart of Methods for PDZ-3$_{\mathrm{Dlg}}$ Modelling (shaded boxes)

6.3 Results and Discussion

A representation of the PDZ-3_{Dlg} structure determined by X-ray crystallography (Cabral *et al.*, 1996) is shown in Figure 6.2. The structure consists of a five-stranded anti-parallel β-barrel flanked by three α-helices.

6.3.1 Determination of Protein-Ligand Interactions

The ligand binding site was determined using SURFNET. The program found only one cavity (volume 720Å^3), shown in Figure 6.3 (same orientation as Figure 6.2), that is large enough for ligand binding (the volume of the second largest cavity was found to be only 150Å^3). Also, over 83% of ligand binding sites are in the largest cavity; and of those that are in another cavity there is little difference in cavity size than ones larger than itself (Laskowski *et al.*, 1996b).

The largest cavity in PDZ-3_{Dlg} includes a highly conserved hydrophobic sequence within the PDZ domain family (*i.e.* Gly-475, Leu-476, Gly-477 and Phe-478) as well as other residues (*i.e.* Ile-480, corresponding to a largely aliphatic sequence position, and Leu-532, corresponding to a largely conserved leucine position), all of which, together, form a hydrophobic pocket. At one end of the hydrophobic pocket is Arg-471, corresponding to a largely conserved basic residue position in the PDZ domain family. The Arg-471 sidechain is held in rigid conformation by hydrogen bonds to three mainchain carbonyl oxygens (*i.e.* those of Thr-474, Leu-532 and Gly-536). Also, a positively charged residue (arginine or lysine) is present in this position in all the PDZ domains that are known to bind C-terminal peptides, which corresponds to approximately 85% of known PDZ domains (Cabral *et al.*,

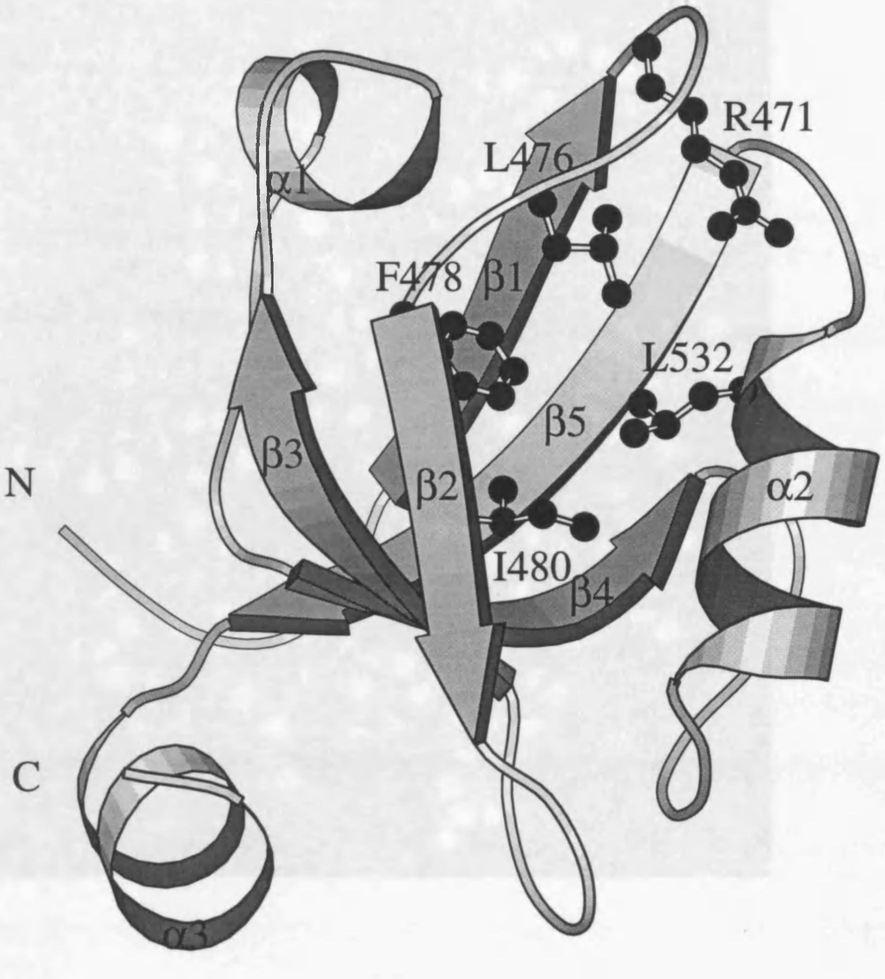


Figure 6.2 Crystal structure of the PDZ-3 domain in ribbon representation with secondary structural elements and the N- and C- termini indicated. Sidechains forming the hydrophobic pocket (residues L476, F478, I480 and L532) along with R471 are also shown.

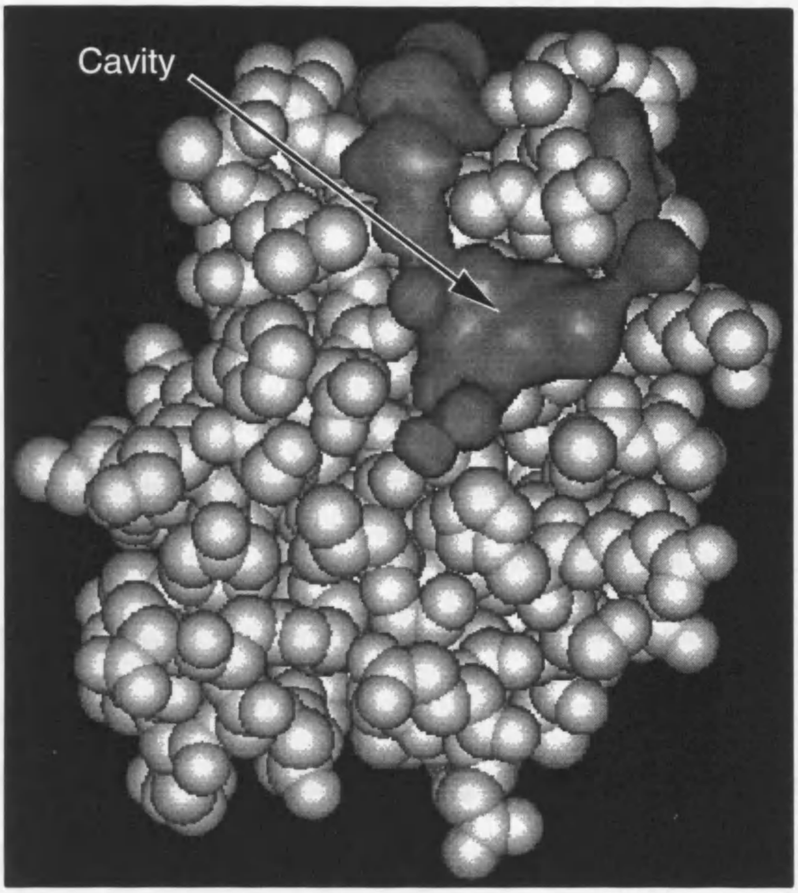


Figure 6.3 Space–filled represenation of PDZ–3$_{Dlg}$ (same orientation as Figure 6.2) illustrating location of largest cavity (shown filled) in the protein.

1996). Figure 6.4 (same orientation as Figure 6.2) illustrates the hydrophobic binding pocket and Arg-471 with respect to the PDZ-3_{Dlg} molecular surface (generated using GRASP, Nicholls *et al.*, 1991; the surface potentials were mapped from formal atomic charges on the atoms within PDZ-3_{Dlg}).

The GRID contour maps were used as a guide to manually dock the Thr-Asp-Val-COO^- ligand into PDZ-3_{Dlg} in order to determine distance constraints for automated modelling of the complex by MODELLER.

Of the possibilities available from the GRID contour maps (as well as the information above), the following relationship between the tri-peptide and the PDZ-3_{Dlg} were proposed: (1) the existence of a salt bridge between the carboxylate group in the ligand and the Arg-471 sidechain in PDZ-3_{Dlg}, (2) hydrophobic interactions between the valine sidechain in the ligand and the hydrophobic pocket in PDZ-3_{Dlg}, (3) the ligand chain as well as the position of the threonine sidechain was deemed to lie along the groove that runs over the top of the molecule in Figure 6.4, (4) the aspartate sidechain in the ligand points away from the molecular surface into solvent, thus rationalising that any residue may occupy this position, and (5) backbone hydrogen bonds were specified between the ligand and PDZ-3_{Dlg}.

Constraints (based on points 1, 2 and 5, above) were set up for the automatic generation of the complex, and are specified in the Tables 6.1 and 6.2 below. The distance between the protein Leu-532 CG and the ligand Val CB was predicted to be lower than those for the other hydrophobic interactions (see Table 6.1) based on structural observations during the manual docking procedure.

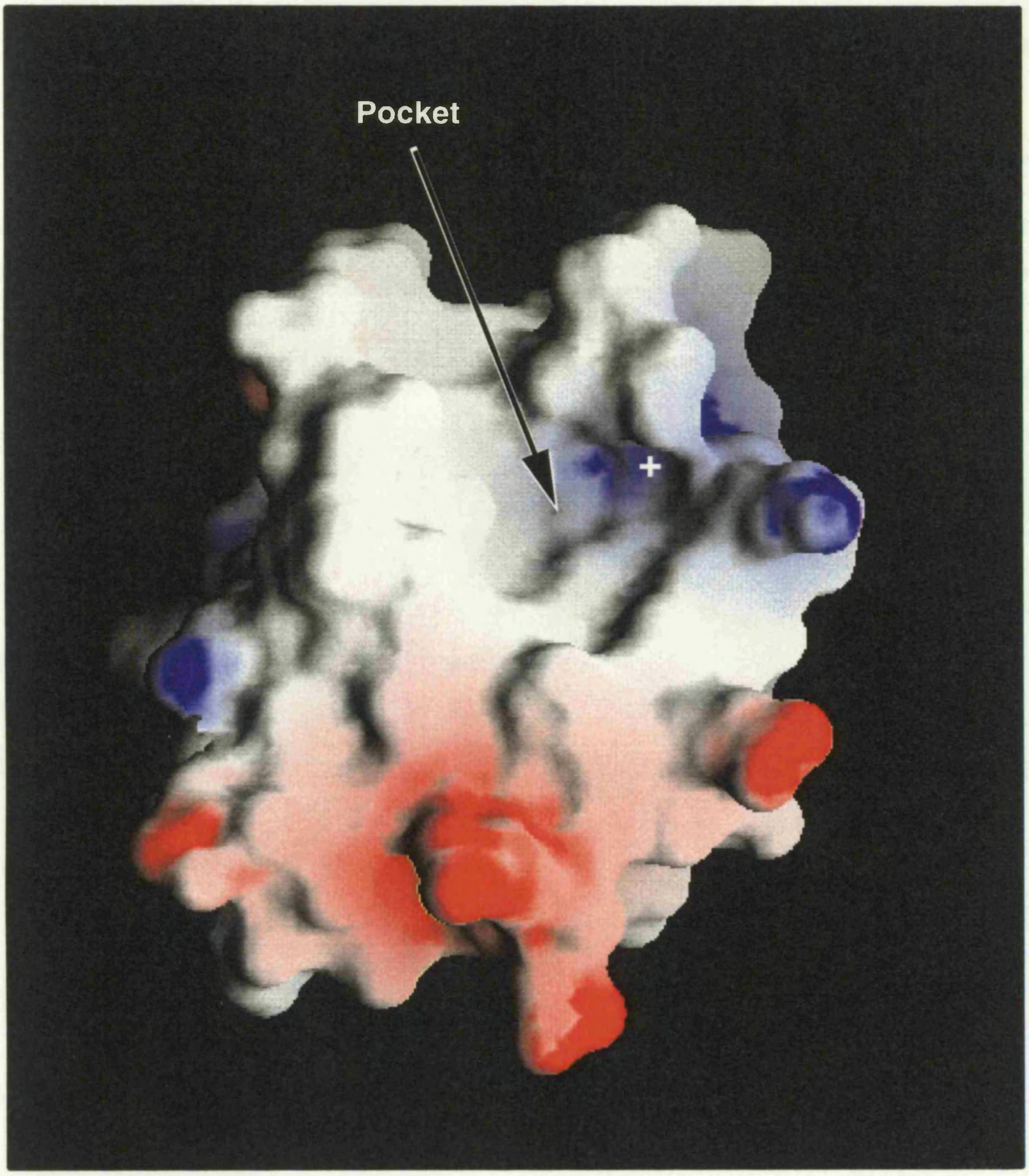


Figure 6.4 Molecular surface representation of PDZ-3$_{Dlg}$ (same orientation as Figure 6.2) illustrating location of hydrophobic pocket as well as R471 (denoted by the white cross)

PDZ-3_{Dlg}	Ligand (Thr-Asp-Val)	Upper Bound (Å)	Lower Bound (Å)
Leu-476 CG	Val CB	9.0	VDW contact
Phe-478 CZ	Val CB	9.0	VDW contact
Ile-480 CB	Val CB	9.0	VDW contact
Leu-532 CG	Val CB	6.5	VDW contact

Table 6.1 Distance constraints for hydrophobic interactions.

PDZ-3_{Dlg}	Ligand (Thr-Asp-Val)	Upper Bound (Å)	Lower Bound (Å)
Arg-471 NH1	Val OXT	3.3	2.5
Phe-478 O	Asp NH	3.3	2.5
Phe-478 NH	Thr OH	3.3	2.5
Gly-477 NH	Thr OH	3.3	2.5

Table 6.2 Distance constraints for hydrogen bonds.

6.3.2 Modelling of Complex

An ensemble of 10 models for the PDZ-3_{Dlg} / ligand complex was generated based on the above distance constraints. The average pairwise RMSD over all non-hydrogen atoms for the homology model ensemble excluding the ligand was 1.00Å. The average pairwise RMSD over all non-hydrogen atoms between the PDZ-3_{Dlg} crystal structure and the homology model ensemble excluding the ligand was 2.95Å.

The structural difference between the homology models and the crystal structure is a global difference rather than being due to any significant local conformational changes in the homology models at or near the binding site; this was confirmed by manual inspection of the structures. However, the homology models are generated using uncomplexed PDZ-3_{Dlg} as the only structural template.

To generate the final model of the PDZ-3_{Dlg} /ligand complex, the 'lowest energy'[44] homology model complex was superposed onto the PDZ-3_{Dlg} crystal structure, using the homology modelled PDZ-3_{Dlg} domain as a reference for superposition. The homology modelled PDZ-3_{Dlg} domain was then discarded, leaving the PDZ-3_{Dlg} crystal structure and the 'bound' ligand. The resultant docking procedure is thus rigid body docking with respect to the uncomplexed form of the PDZ-3_{Dlg} domain.

Figure 6.5 (same orientation as Figure 6.2) illustrates the position of the Thr-Asp-Val-COO^- tri-peptide within PDZ-3_{Dlg} in the final model.

6.3.3 Comparison of Model with Crystal Structure Complex

The PDZ domains of brain post-synaptic protein PSD-95 interact with the C-terminal subunits of K^+ channels. Crystal structures of the third PDZ domain from rat PSD-95 (PDZ-$3_{\mathrm{PSD\text{-}95}}$; PDB code s031, but coordinates not yet available), both free-form and complexed with a nine residue peptide (*i.e.* Thr-Lys-Asn-Tyr-Lys-Gln-Thr-Ser-Val), were solved by Doyle *et al.* (1996). The four C-terminal residues of the peptide in the complex (*i.e.* -Gln-Thr-Ser-Val), corresponding to the motif X-Thr/Ser-X-Val-COO^-, engage directly in ligand

[44] Some of the structures within the ensemble contained unlikely orientations between protein and ligand within part of the structure (visual inspection), possibly due to a lack of user-specified sufficient distance constraints. Therefore, selecting a representative structure based on cluster analysis was not appropriate.

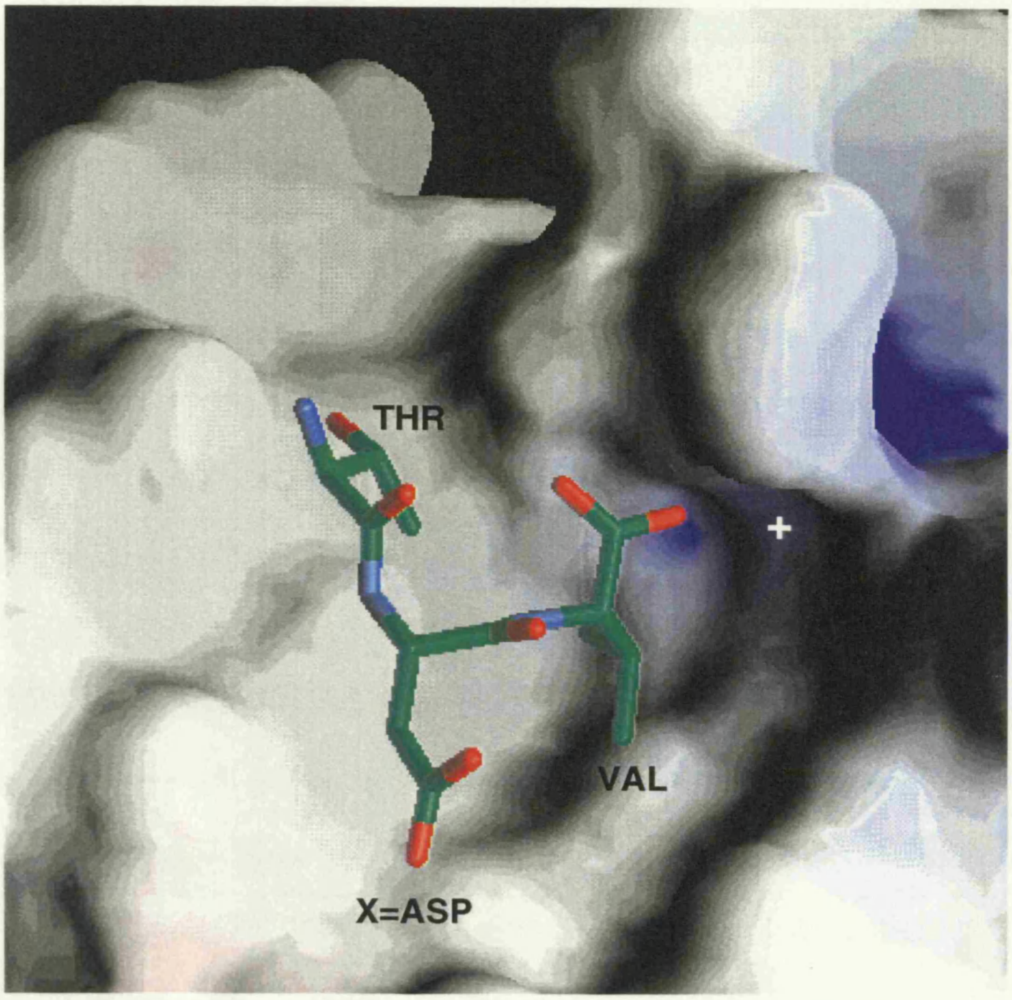


Figure 6.5 Model of tri-peptide ligand, shown using stick representation, bound to PDZ-3$_{Dlg}$, shown using molecular surface representation (same orientation as in Figure 6.2). The white cross denotes the location of R471.

binding. Details of this work were unavailable at the time of the modelling study of PDZ-3_{Dlg} presented here.

In the PDZ-$3_{\mathrm{PSD\text{-}95}}$ complex, the ligand lines the channel corresponding to that between the $\alpha2$ helix and the $\beta2$ strand of PDZ-3_{Dlg} in Figure 6.2, which runs down from the hydrophobic pocket corresponding to that opposite the Arg-471 of PDZ-3_{Dlg} (seen in Figure 6.4), rather than running over the top of the domain.

Doyle *et al.* (1996) report little conformational change associated with peptide binding in PDZ-$3_{\mathrm{PSD\text{-}95}}$ (the Cα RMSD between complexed and uncomplexed PDZ-$3_{\mathrm{PSD\text{-}95}}$ is 0.9Å). Doyle *et al.* (1996) report significant Cα displacements occur between the two PDZ-$3_{\mathrm{PSD\text{-}95}}$ models (approximately 4Å) at Gly-319, Ser-320 and Thr-321 (corresponding to Gly-472, Ser-473 and Thr-474 in PDZ-3_{Dlg}) within the carboxylate binding loop (*i.e.* the loop containing the largely conserved positive residue in PDZ domains, corresponding to Arg-471 in PDZ-3_{Dlg}). However, the displaced residues are located at a crystal contact, and their re-positioning has only a slight effect on the positions of the ligand-interacting atoms; therefore, the conformational change in these residues being attributable specifically to peptide-binding is considered by Doyle *et al.* (1996) to be questionable.

'[O]nly very minor' positional changes in the sidechains of functionally important residues were observed by Doyle *et al.* (1996). However, one notable re-positioning is the His-372 sidechain in PDZ-$3_{\mathrm{PSD\text{-}95}}$ (corresponding to His-525 in PDZ-3_{Dlg}) which, in uncomplexed form, is hydrogen bonded *via* the histidine imidazole ring to the backbone carbonyl of Ile-327 (Ile-480 in PDZ-3_{Dlg}); when complexed with the ligand, the histidine sidechain undergoes a 180° rotation in order to hydrogen bond to the sidechain oxygen of Thr -2 in the ligand (*i.e.* -Gln-*Thr*-Ser-Val-COO$^-$), and also forms a hydrogen bond to the backbone carbonyl of Gly-329 (Gly-482 in PDZ-3$_{\mathrm{Dlg}}$), both *via* the histidine's imidazole ring (Doyle *et al.*, 1996). The hydrogen bond between His-525 (in PDZ-3$_{\mathrm{Dlg}}$) and the threonine sidechain (in Thr-Asp-Val-COO$^-$) was not specified in the constraints list for the modelling study; this omission arises because the conformational change in the sidechain of His-525 was not predicted (there was no previous evidence for such a conformational change).

Doyle *et al.* (1996) report hydrophobic sidechain interactions between the C-terminal valine in the nine residue ligand and residues Leu-323, Phe-325, Ile-327 and Leu-379, which line the 'hydrophobic pocket' in the PDZ-3$_{\mathrm{PSD\text{-}95}}$ domain. These residues in PDZ-3$_{\mathrm{PSD\text{-}95}}$ correspond to residues Leu-476, Phe-478, Ile-480 and Leu-532 in PDZ-3$_{\mathrm{Dlg}}$ for which distance constraints were specified for the modelled PDZ-3$_{\mathrm{Dlg}}$ complex model.

Also, the sidechain of the residue preceding the ligand's C-terminal valine points out into solution in the PDZ-3$_{\mathrm{PSD\text{-}95}}$ complex, as does that in the PDZ-3$_{\mathrm{Dlg}}$ complex model.

A very significant difference between the two PDZ-3$_{\mathrm{PSD\text{-}95}}$ X-ray structures is the presence of a well-ordered water molecule positioned between, and hydrogen bonded to, the ligand's terminal carboxylate group and the guanidinium group of Arg-318 (Arg-471 in PDZ-3$_{\mathrm{Dlg}}$) within the complex. In the uncomplexed form the water molecule is shifted in its position and is less well defined (Doyle *et al.*, 1996). It may be assumed, on the basis of the homology between these two PDZ domains, that the structural involvement of this water molecule in complex formation occurs in PDZ-3$_{\mathrm{Dlg}}$ as well as PDZ-3$_{\mathrm{PSD\text{-}95}}$. This would have a significant effect on the modelling of the PDZ-3$_{\mathrm{Dlg}}$/tri-peptide complex since the tri-peptide would thus be directed towards α2-β2 channel rather than the channel over the top of PDZ-3_{Dlg}, leading to an amended set of distance constraints for the model generation.

The carboxylate group of the ligand also forms hydrogen bonds with the backbone nitrogens of three residues within the Gly-Leu-Gly-Phe motif of PDZ-$3_{\mathrm{PSD\text{-}95}}$, namely Leu-323, Gly-324 and Phe-325 (corresponding to Leu-476, Gly-477 and Phe-478 in PDZ-3_{Dlg}). None of these hydrogen bonds occur in the model of the PDZ-3_{Dlg} complex since the modelled ligand was mis-positioned due to (1) the absence of information regarding the existence of the mediating water molecule between Arg-471 of PDZ-3_{Dlg} and the carboxylate terminus of the ligand, and (2) the failure to predict the repositioning of the His-525 sidechain.

Finally, in PDZ-$3_{\mathrm{PSD\text{-}95}}$, the backbone of the ligand forms hydrogen bonds with the backbone of two of the hydrophobic pocket residues. That is, C-terminal valine amide nitrogen to Phe-325 (Phe-478 in PDZ-3_{Dlg}) carbonyl oxygen, Thr -2 carbonyl oxygen to Ile-327 (Ile-480 in PDZ-3_{Dlg}) amide nitrogen, and Thr -2 amide nitrogen to Ile-327 (Ile-480 in PDZ-3_{Dlg}) carbonyl oxygen (*i.e.* similar to an anti-parallel β-strand/β-strand interaction). None of these hydrogen bonds were predicted, and were therefore not specified as constraints for the PDZ-3_{Dlg} complex model. The PDZ-$3_{\mathrm{PSD\text{-}95}}$ hydrogen bonds relate to a relative positional shift of the ligand away from Arg-318 (Arg-471 in PDZ-3_{Dlg}) compared with that in the PDZ-3_{Dlg} model, which is positioned closer to the arginine. Again, this is directly attributable to the lack of information regarding the water molecule mediating the interaction between the Arg-471 guanidinium group in PDZ-3_{Dlg} and the C-terminal carboxylate group in the ligand.

6.4 Conclusion

Modelling of the protein-ligand complex was based on predicted structural details including (1) the interaction between the ligand C-terminal carboxylate group and Arg-471 in PDZ-3_{Dlg}, (2) the interaction between the ligand C-terminal valine sidechain and protein sidechains in a hydrophobic pocket within the predicted binding cavity, and (3) hydrogen bonds between the tri-peptide ligand and the protein.

Subsequent determination of the crystal structure of a homologous complex revealed that the result of the modelling was not completely correct. Nevertheless, a particularly significant piece of information could not be predicted: the presence of a mediating water molecule between the carboxylate group in the ligand and the basic sidechain corresponding to that of Arg-471 in PDZ-3_{Dlg}. Knowledge of this could possibly have directed the tri-peptide towards the correct channel, which in turn could perhaps have lead to the identification of a hydrogen bond interaction between the sidechain oxygen of Thr -2 in the ligand and His-525 in PDZ-3_{Dlg}.

CHAPTER 7

ELECTRON TRANSFER AND SUBSTRATE SPECIFICITY IN TRIMETHYLAMINE DEHYDROGENASE (TMADH)

7.1 Introduction

Bacterial trimethylamine dehydrogenase (TMADH) is a homodimeric enzyme with an interfacial surface area of approximately 6000Å^2. Each monomer consists of 729 residues, a flavin mononucleotide (FMN) moiety covalently bound to Cys-30 SG at the C6 position in the isoalloxazine ring, an iron-sulphur (4Fe-4S) cluster (each iron being covalently bound to one of Cys-345, Cys-348, Cys-351 and Cys-364 SG atoms), and an adenosine diphosphate (ADP) moiety (the ADP is not involved in the catalysis and is thought to be a component of the redundant remains of an ancestral ADP-binding domain, now used only as part of the 'scaffolding' for the TMADH structure).

TMADH is involved in electron transfer from trimethylamine (TMA) to the FAD- (flavin adenine dinucleotide) containing target protein, 'electron transfer flavoprotein' (ETF), through the following reaction:

$$(CH_3)_3N + H_2O \text{ -> } (CH_3)_2NH + CH_2O + 2e^- + 2H^+$$

The electron transfer within the enzyme occurs firstly *via* the FMN group then *via* the 4Fe-4S cluster. The electron is then thought to travel to Tyr-442 at the surface of the enzyme (*i.e.* the closest surface residue to the 4Fe-4S cluster), a spatial distance of 11.6Å (Wilson *et* *al.*, 1995). The Tyr-442 residue is located within a concave region on the enzyme surface, covering an area of approximately 1200Å^2, and this region may be the site of interaction with ETF (Wilson *et al.*, 1995). Mutations of Tyr-442 exhibit decreased catalytic activity of the enzyme compared to the wild type (Wilson *et al.*, 1997). A structural study of the enzyme with respect to electron transfer could serve to rationalise the pathway of the electrons during the catalytic process.

The substrate binding site in TMADH consists of the isoalloxazine ring in FMN on one side of the pocket and, on the other, an 'aromatic bowl' consisting of Tyr-60, Trp-264 and Trp-355, with each sidechain binding one of the three methyl groups in TMA *via* a cation-π interaction. The location of this binding site is buried within the interior of the enzyme, but is close to the 2-fold symmetry axis of the TMADH homodimer.

The closely related enzyme dimethylamine dehydrogenase (DMADH) has a sequence identity of approximately 64% with TMADH (DMADH contains a 6 residue insertion with respect to TMADH). The DMADH substrate, dimethylamine (DMA), undergoes a demethylation during the catalysis, as does TMA with TMADH. The residues corresponding to the active site in these enzymes are highly conserved. Basran *et al.* (1997) have identified three residues in the substrate binding region of TMADH (*i.e.* Tyr-60, the only non-conserved 'aromatic bowl' residue between TMADH and DMADH, and also Ser-74 and Trp-105) as probable determinants for TMA specificity, and mutated these residues to the corresponding amino acids in DMADH in order to test for any resulting specificity for DMA over TMA. The only significant effect on activity was attributed to the Y60Q mutation with TMA as substrate, where mutant enzyme activity decreased significantly compared with wild type TMADH. However, an insight into the mechanism of DMA specificity remains elusive. Further study in substrate specificity using structural rationale to help identify residues instrumental in DMA specificity, would increase understanding in this aspect of functionality.

The crystal structure of TMADH has been solved at 2.4Å resolution (PDB code 2TMD; Lim *et al.*, 1986). However, a recently refined structure at 1.8Å resolution (S.A. White & F.S. Mathews, St Louis, USA, unpublished results) was available for the studies presented here. No experimentally-derived structure exists for DMADH.

The objective of this modelling study was to examine the 3D structure of TMADH and electrostatic surface of TMADH in order to investigate the structural rationale behind TMADH function. An attempt was made to determine the electron transfer pathway from the FMN group, *via* the 4Fe-4S cluster, to Tyr-442. Additionally, substrate access to the binding site of TMADH was investigated. Finally, a structural study was undertaken in order to identify key residues with respect to the enzyme's specificity to TMA over DMA. All of the work was performed in collaboration with Dr. N.S. Scrutton and co-workers (Department of Biochemistry, University of Leicester, UK).

7.2 Methods

The methodological approach to this study is described by the flowchart in Figure 7.1. The methods are detailed in chapter 2.

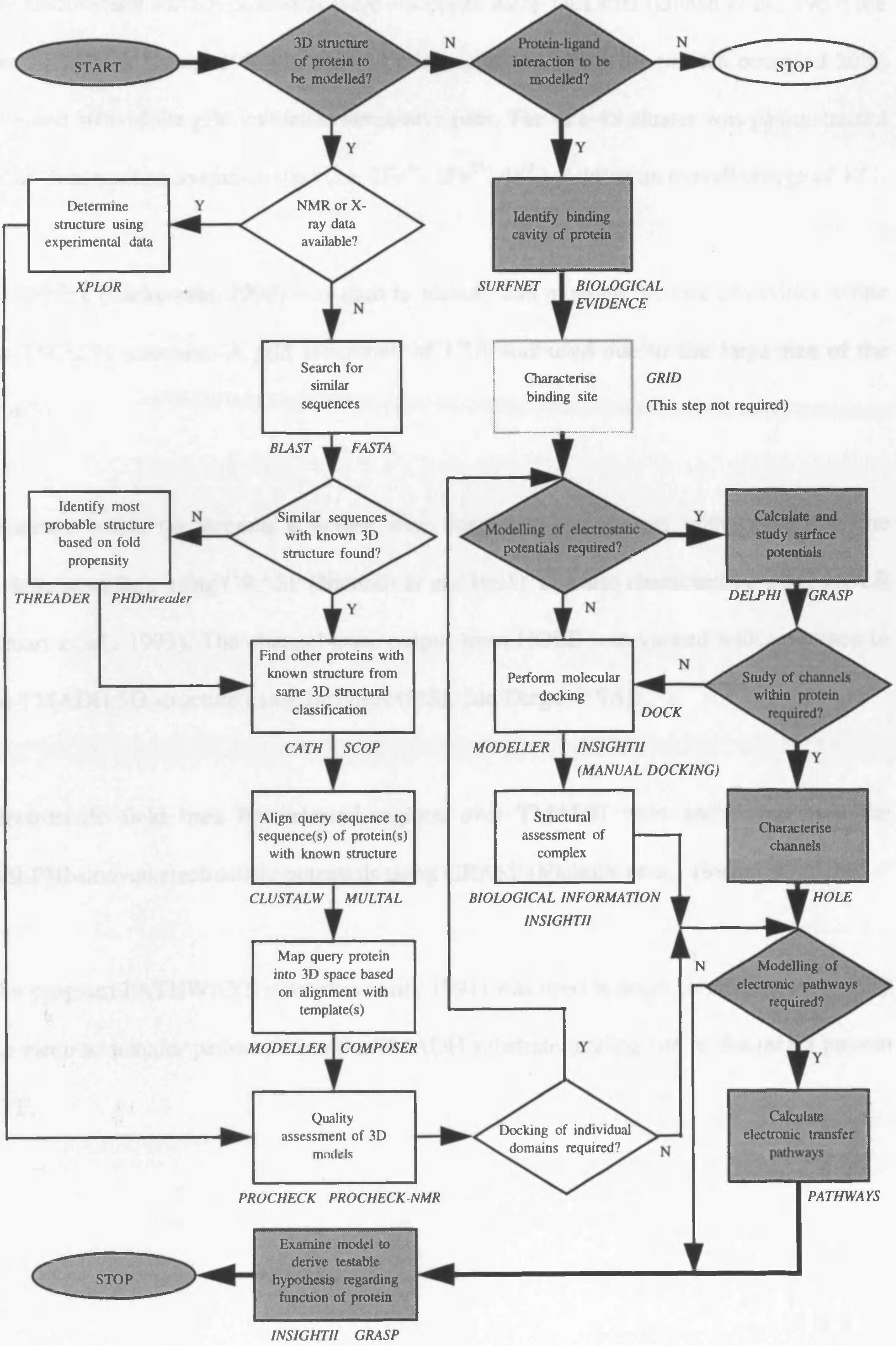

START
3D structure of protein to be modelled?
N
Protein-ligand interaction to be modelled?
N
STOP
Y
Y
Determine structure using experimental data
Y
NMR or X-ray data available?
Identify binding cavity of protein
XPLOR
SURFNET
BIOLOGICAL EVIDENCE
N
Search for similar sequences
Characterise binding site
GRID
(This step not required)
BLAST
FASTA
Identify most probable structure based on fold propensity
N
Similar sequences with known 3D structure found?
Modelling of electrostatic potentials required?
Y
Calculate and study surface potentials
THREADER
PHDthreader
Y
N
DELPHI
GRASP
Find other proteins with known structure from same 3D structural classification
Perform molecular docking
N
Study of channels within protein required?
DOCK
CATH
SCOP
MODELLER
INSIGHTII
(MANUAL DOCKING)
Y
Align query sequence to sequence(s) of protein(s) with known structure
Structural assessment of complex
Characterise channels
CLUSTALW
MULTAL
BIOLOGICAL INFORMATION
INSIGHTII
HOLE
Map query protein into 3D space based on alignment with template(s)
N
Modelling of electronic pathways required?
MODELLER
COMPOSER
Y
Y
Quality assessment of 3D models
Docking of individual domains required?
N
Calculate electronic transfer pathways
PROCHECK
PROCHECK-NMR
PATHWAYS
STOP
Examine model to derive testable hypothesis regarding function of protein
INSIGHTII
GRASP

Figure 7.1 Flowchart of Methods for TMADH Modelling (shaded boxes)

The electrostatic surface potentials were calculated using DELPHI (Gilson *et al.*, 1987; see also Nicholls & Honig, 1991) using a focusing approach where the protein occupied 20%, 50% and 80% of the grid volume in successive runs. The 4Fe-4S cluster was parameterised for an intermediate oxidation state (*i.e.* $2Fe^{3+}$, $2Fe^{2+}$, $4S^{2-}$) yielding an overall charge of +2.

SURFNET (Laskowski, 1995) was used to identify and measure the size of cavities within the TMADH structure. A grid separation of 1.3Å was used due to the large size of the protein.

Channels within the protein structure were identified and viewed with respect to the molecular surface using GRASP (Nicholls *et al.*, 1991), and also characterised using HOLE (Smart *et al.*, 1993). The channel-trace output from HOLE was viewed with reference to the TMADH 3D structure using InsightII (MSI, San Diego, USA).

Electrostatic field lines for selected regions over TMADH were calculated from the DELPHI-derived electrostatic potentials using GRASP (Nicholls *et al.*, 1991).

The program PATHWAYS (Beratan *et al.*, 1991) was used in order to attempt to identify the electron transfer pathway from the TMADH substrate binding site to the target protein ETF.

7.3 Results and Discussion

7.3.1 Electron Transfer Pathway

For the electron transfer pathway calculations, the FMN moiety was used as the reference point for the TMADH substrate binding site, and Tyr-442 (the nearest surface residue to the 4Fe-4S cluster) was used as the reference point for the target protein, ETF.

The first calculation specified both these endpoints, that is, to calculate the optimum pathway from the FMN's isoalloxazine ring (bond C7-C8) to Tyr-442 (bond CZ-OH). The result suggested that Arg-322 (bond NE-HE), Arg-299 (bonds CG-HG1 and NE-HE) and Glu-439 (OE2 and one of its lone pair of electrons) were associated with the pathway. However, the resulting pathway omitted the 4Fe-4S cluster (electrons are known to pass through this cluster), although its iron and sulphur atoms were parameterised. This result therefore indicates an over-simplistic treatment of electronic orbitals by the program. Arg-322, Arg-299 and Glu-439 may not therefore be associated with the electron transfer pathway, especially since the through-space jump from FMN to Arg-322 occurred *via* a non-isoalloxazine ring oxygen; however, Glu-439 is identified again in a subsequent calculation below.

To overcome the problem of an incorrectly modelled pathway, PATHWAYS was used to calculate an electron pathway from the FMN's isoalloxazine ring to the 4Fe-4S cluster, and then to calculate a subsequent pathway from the 4Fe-4S cluster to Tyr-442.

For the FMN isoalloxazine ring (bond C7-C8) to the 4Fe-4S cluster (bond FE3-S3) calculation, the result suggested that Cys-351 (SG and its associated lone pair of electrons) was associated with the pathway. The through-space jump from FMN occurred *via* C7M-H7M1 bond in the isoalloxazine ring (an 'acceptable' electron transition point) to Cys-351. Cys-351 is directly associated with the 4Fe-4S cluster, since the Cys-351 SG atom is covalently bound to the FE3 iron atom.

For the 4Fe-4S cluster (bond FE1-S1) to Tyr-442 (bond CZ-OH), the result suggested that Cys-345 (SG and its associated lone pair of electrons) and Glu-439 (OE2 and one of its lone pair of electrons) were associated with the pathway. Cys-345 is directly associated with the 4Fe-4S cluster, since the Cys-345 SG atom is covalently bound to the FE1 iron atom.

From the above results, other than cysteine residues directly associated with the 4Fe-4S cluster, Glu-439 seems to be implicated in electron transfer from the substrate (and particularly from the 4Fe-4S cluster) to Tyr-442. However, manual inspection of the 3D structure of TMADH indicates that His-443, positioned opposite the 4Fe-4S cluster, may possibly be another candidate for involvement in electron transfer from the 4Fe-4S cluster to Tyr-442. Also from manual inspection, Wilson *et al.* (1997) identified Val-344 (located at the surface close to Tyr-442) as an alternative pathway residue to the ETF, rather than the PATHWAYS-derived Glu-439 and Tyr-442; both Val-344 and Glu-439 are in close spatial proximity to both Cys-345 (which is directly associated with the 4Fe-4S cluster) and Tyr-442.

7.3.2 Role of C-terminal Residues

Another focus of the work on TMADH with the collaborators was the role of the C-terminal residues in enzyme function. Residues 713 to 729 from each monomer occur on the surface of the protein dimer, embracing residues on the other sub-unit. One hypothesis for the function of these residues was that they could affect the relative proximity of the monomers and therefore, perhaps, influence substrate access to the (buried) binding site.

The 3D structure of TMADH was examined, along with the modelling of the electrostatic potentials over the enzyme surface using DELPHI for both wild type and mutant enzymes (the mutations involving sequential deletions in the residue range 713-729); a representation of the 17 C-terminal residues with respect to the molecular surface is illustrated in Figure 7.2. The structural examination concluded that the effect of these 17 C-terminal residues did not appear to be involved in substrate access, nor did they seem to have any significant involvement in the function of the enzyme other than contributing to the stability of the dimer.

The collaborators have recently undertaken experimental studies relating to the 17 C-terminal residues in TMADH (Ertughrul *et al.*, in press). They found that the deletion of these residues (even that of the 5 C-terminal residues) strongly diminishes, but does not completely remove, the ability of the enzyme to covalently bind the FMN moiety (located over 20Å away from the C-terminus). They also find that deletion of these residues does not significantly affect either dimer stability or the overall structural integrity of the enzyme; however, minor conformational changes in the mutant enzymes are reported to occur since the hydrophobic exposures on progressive C-terminal deletions were not consistent with

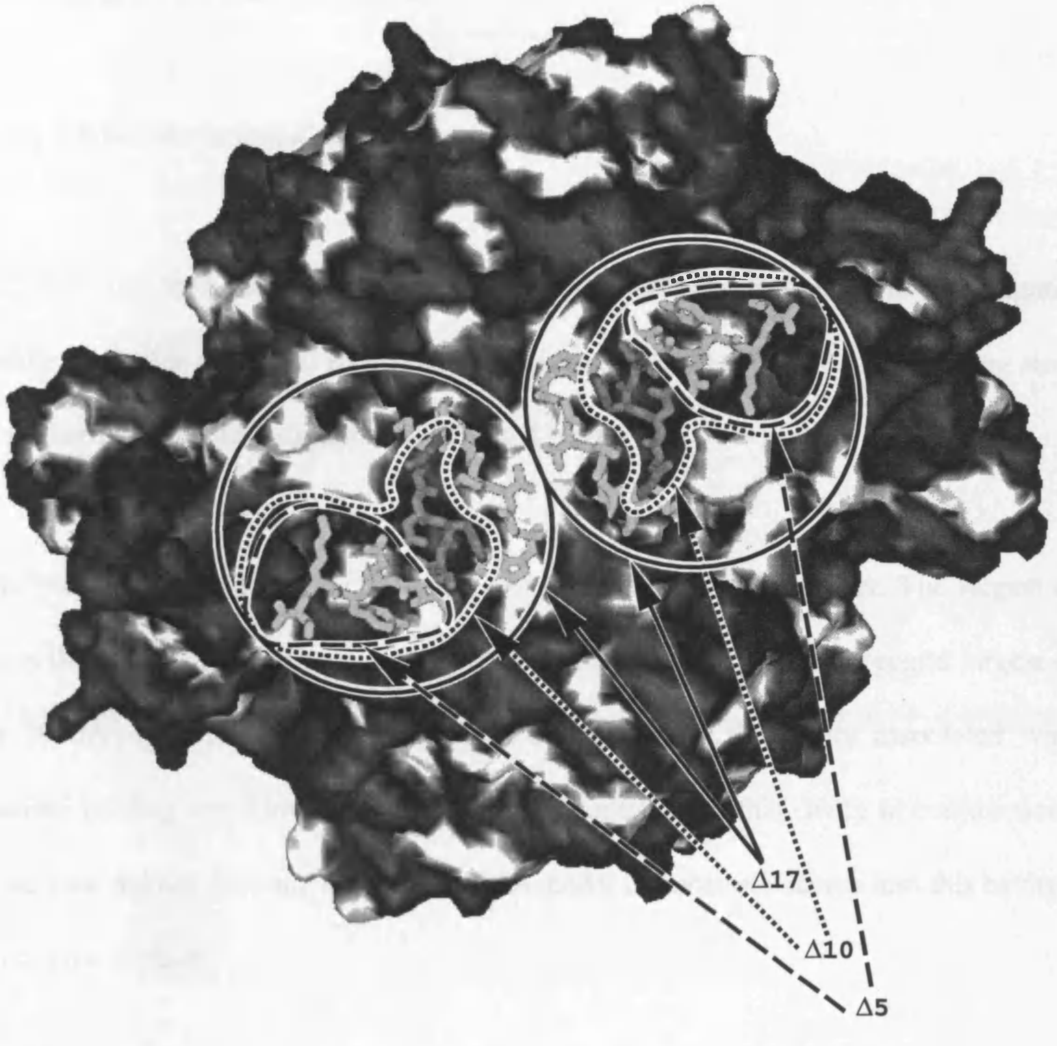


Figure 7.2 The 17 C–terminal residues (stick) of TMADH shown in conjunction with the molecular surface of the remainder of the TMADH dimer. The solid circles describe the regions containing the 17 C–terminal residues (Δ17), the dotted perimeters describe the regions containing the 10 C–terminal residues (Δ10), and the dashed perimeters describe the regions containing the 5 C–terminal residues (Δ5).

expectations. It is possible that the minor conformational changes occurring from the absence of the C-terminal residues in the mutants have an indirect structural effect on the substrate binding site, causing a reduction in the ability of the enzyme to covalently bind FMN during the folding of TMADH.

7.3.3 Characterisation of Channels

The modelling study of the 17 C-terminal residues (above) stimulated an investigation to identify a possible structural rationale for substrate access to the TMADH binding site (one per monomer). This investigation is presented below.

SURFNET was used to characterise cavities within the TMADH dimer. The largest cavity was in the centre of the protein dimer, with a volume of 31000Å^3; the second largest cavity was 1900Å^3. The largest cavity was thus deemed to be the cavity associated with the substrate binding site. However, graphical representation of this cavity in conjunction with the enzyme did not give any indication of the model of substrate access into this cavity from the enzyme surface.

The presence of three inter-connecting channels were identified within the protein using a molecular surface representation of TMADH within GRASP. One of the channels (see Figure 7.3) lies at the dimer interface and runs along the 2-fold symmetry axis within the dimer. The other two channels (see Figure 7.4) are non-central, but related by two-fold symmetry, and connect with the central channel. The existence of the enzyme as a dimer gives rise to the formation of these channels; none of the channels exists within a

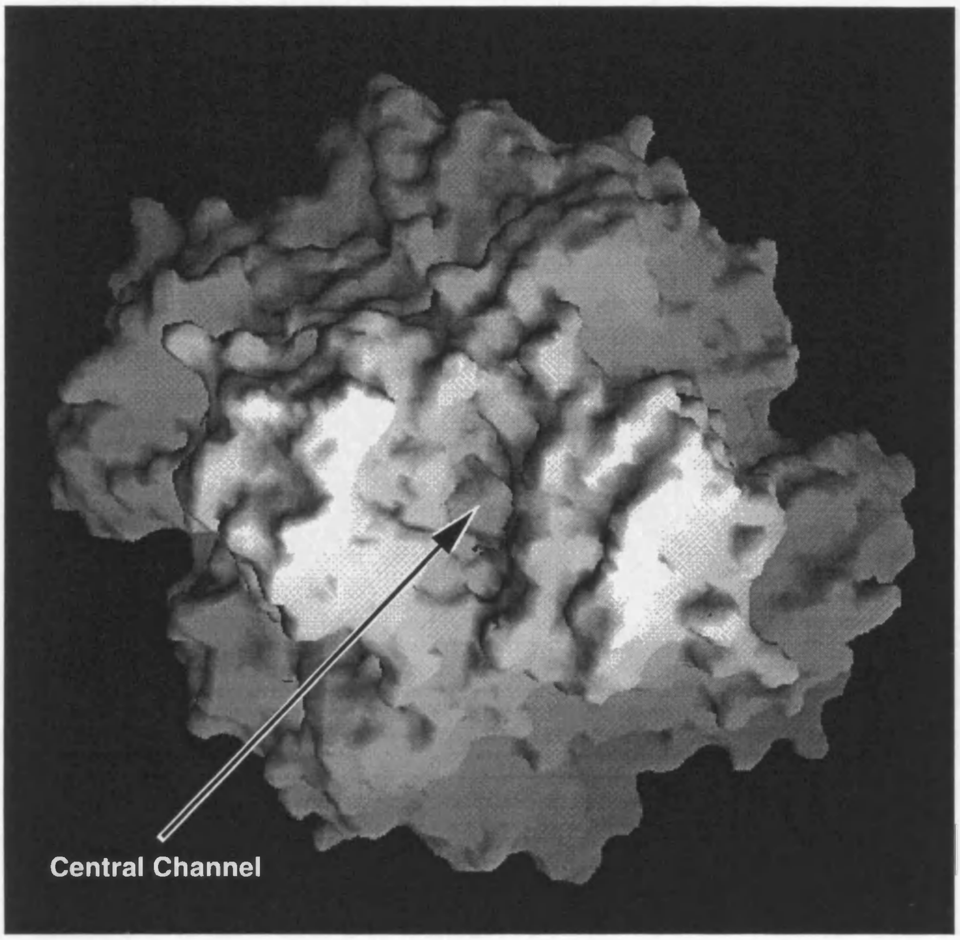


Figure 7.3 Channel running along two–fold axis and leading into a large cavity within the TMADH dimer. Entrance to channel (above) is on opposite end of enzyme to the C–terminal residues.

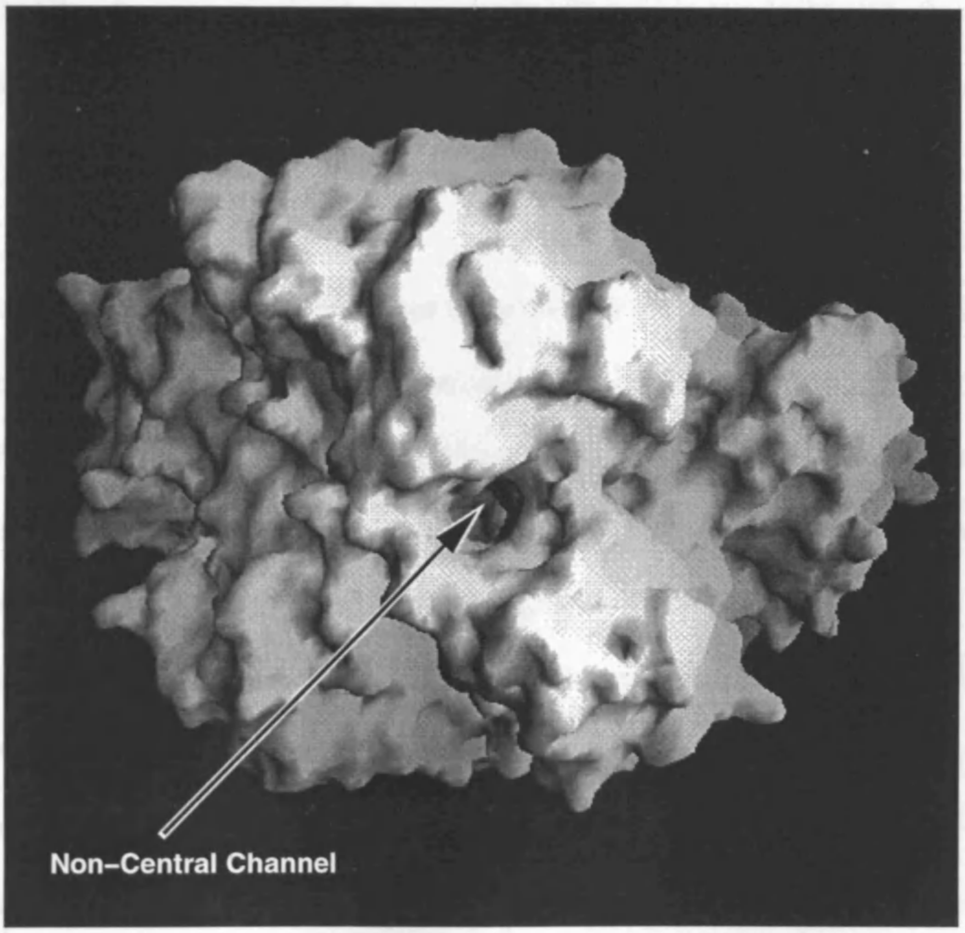


Figure 7.4 Location of one of two non–central channels which are symmetric about the two–fold axis, and lead into a large cavity within the TMADH dimer.

monomeric subunit alone. Furthermore, the existence of the two channels away from the two-fold axis had not been reported previously.

All the channels have direct access to the centre of the protein, from where there is easy access to the TMA binding site (*i.e.* the 'aromatic bowl' opposite the FMN) of both chains. His-71 and Phe-570 were identified as being positioned approximately at the centre of the protein dimer.

The paths of the channels were characterised using the program HOLE. The central channel has a minimum radius of 1.5Å. This minimum radius is located close to His-563 (from both chains), where the channel becomes relatively constricted. Other residues lining the central channel include Asn-566, His-569 and Asp-591 (all from both chains). The non-central channels have a minimum radius of 3.0Å, and, unlike the central channel, do not seem to be constricted in any particular region. Residues among those lining the non-central channels are His-569 and Phe-570 (both are from a single chain and line the associated non-central channel), and His-563, Asn-566 and Asp-591 (all from both chains). The association of residues His-563, His-569, Asn-566 and Asp-591 with both central and non-central channels is due to the fact that all the channels meet within the enzyme.

The molecular surface of the channels were isolated by scribing around them within GRASP and the electrostatic potentials, calculated using DELPHI, were also read in for the channel regions. Then, electrostatic field lines, seeded at these channel regions of the molecular surface and user-specified as travelling from a negative to positive direction, were calculated within GRASP. The field lines would thus convey the movement of the substrate, TMA (protonated, and therefore positively charged, in solution), from bulk solvent towards the protein surface. The results suggested substrate access into the binding site *via* the non-central channels and not the central channel (the diameter of the non-central channels will readily accommodate the access of TMA but the central channel will not). Field lines, seeded from the channels, that were user-specified as going from a positive to negative direction travelled back into the surface, conveying opposition to access for a negatively charged moiety. His-563, which lines the smaller, central channel may possibly play a key role in removal of the reaction products.

7.3.4 Substrate Specificity

The mutation studies reported by Basran *et al.* (1997) relating to TMA/DMA specificity in TMADH(/DMADH) involve the following mutations in TMADH: Y60Q, S74T and W105F. The collaborators of this present study also have data involving the mutation T257N in TMADH (unpublished results). The four mutants are Y60Q, Y60Q/S74T, Y60Q/S74T/W105F and Y60Q/S74T/W105F/T257N.

The effect of the Y60Q mutation in the 'aromatic bowl' of TMADH is to remove cation-π bonding to one of the three methyl groups in TMA, thereby decreasing specificity for TMA. It is also most likely to lead to the formation of a hydrogen bond between the N-H hydrogen in DMA and Q60 OE1, thereby increasing specificity for DMA.

The further mutations, S74T and W105F, are more subtle with respect to TMA/DMA specificity. Both residues are in close spatial proximity to the 'aromatic bowl', and their involvement in substrate specificity, while not completely clear, is most likely to be attributable to direct and indirect steric effects due to changes in sidechain bulk (the W105F mutation also decreases aromatic surface area).

The fourth mutation, T257N, is in a spatially distinct region of the structure to the other three mutations. It appears to remove a hydrogen bond between T257 OG and a crystallographic water (34H), replacing it with a hydrogen bond interaction between N257 ND2 and O3' of the FMN group. This may lead to a rearrangement in FMN conformation, although this does not seem to have a drastic effect on TMA/DMA specificity.

The activity of wild type TMADH is more than 20-fold lower for DMA than for TMA as substrate (Basran *et al.*, 1997). Catalytic activity decreased significantly from wild type TMADH for the Y60Q mutant using TMA as substrate, with relatively little subsequent change for the remaining mutants. No significant changes in activity occurred for the mutants (including the Y60Q mutant) compared with wild type TMADH when using DMA as substrate.

The above mutations do not appear to sufficiently rationalise TMA/DMA specificity. Therefore, the crystal structure of TMADH was manually inspected in order to select further residues for mutation studies of TMADH for the purpose of achieving DMA-specificity over TMA-specificity. This selection, and the reasoning behind it, is as follows.

In the TMADH wild type, Tyr-60 (an 'aromatic bowl' residue) appears to be held in position by a hydrogen bonding network involving Asp-69, Arg-72 and Glu-131 (see Figure 7.5). In DMADH, the equivalent residue to Glu-131 in TMADH is an alanine, therefore the salt bridge between Arg-72 and Glu-131 present in TMADH cannot be present in the

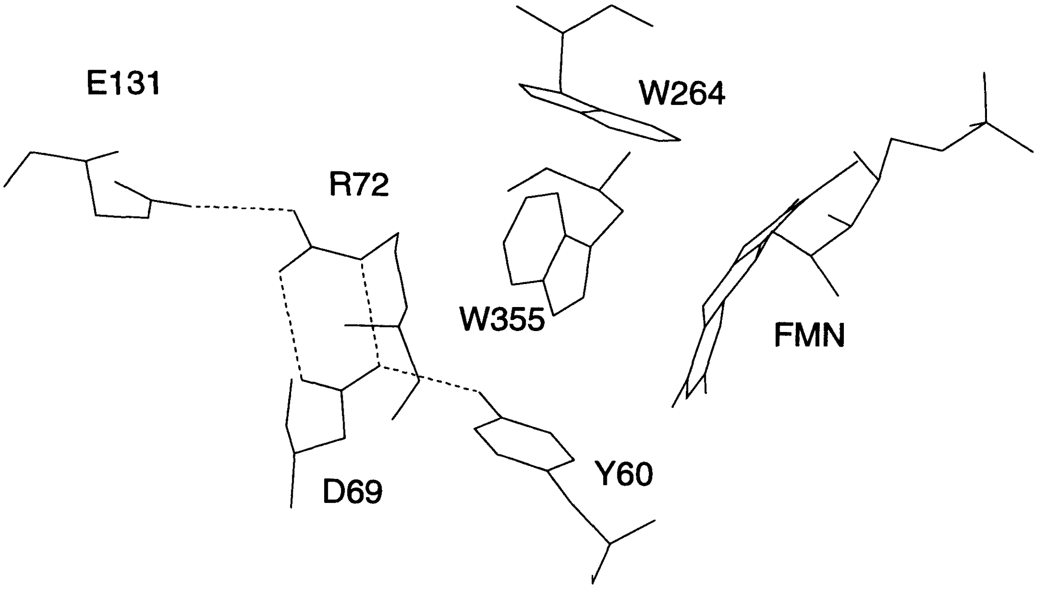


Figure 7.5 The hydrogen bond (dashed lines) network involved in stabilising the sidechain of Y60, one of the three 'aromatic bowl' residues in TMADH.

equivalent position in DMADH. Both Asp-69 and Arg-72 are conserved in TMADH and DMADH. However, for this region of the sequence, residues His-71, Leu-73 and Ser-74 in TMADH are in equivalent positions to Leu, Ile and Thr, respectively, in DMADH. These non-conserved residues, due to their structural association, were selected as candidates for mutation in addition to the previously mutated residues.

The newly-proposed TMADH mutants, based on the observed hydrogen bond network, are:

(1) the penta-mutant Y60Q/S74T/W105F/T257N/E131A

(2) the hexa-mutant Y60Q/S74T/W105F/T257N/H71L/L73I

(3) the hepta-mutant Y60Q/S74T/W105F/T257N/H71L/L73I/E131A

Experiments regarding the kinetic and thermodynamic activity of these mutants were performed by Dr. N. Scrutton and co-workers (Department of Biochemistry, University of Leicester, UK). It was only possible to express the first two of these mutant enzymes. Within these enzymes, structural perturbations occurred such that the FMN prosthetic group was not properly incorporated into the substrate binding site. The mutant enzymes become inactivated after one substrate reaction. The role of the hydrogen bonding network residues were therefore not fully rationalised.

Aside from this set of three mutations, another mutation was proposed (the 'insertion mutant') based on different criteria, explained as follows. A 6-residue insertion in DMADH is located between T132 and L133 of TMADH, and is the only insertion/deletion between the two proteins.

i.e.	DMADH	$_{130}$FATVPGCPGFTY$_{141}$
	TMADH	$_{130}$FET------LSY$_{135}$

T132 is, obviously, spatially close to E131 which is a residue that is part of the hydrogen bonding network associated with the substrate binding site (see Figure 7.5) and which a residue involved in the hepta-mutation specified above. A manual inspection of the 3D structure of TMADH showed that the position of the DMADH (no expermintal 3D structure) insertion would be structurally well placed to restrict substrate access to the FMN group and the associated 'aromatic bowl', thereby affecting substrate specificity most probably with respect to substrate size. A suitable mutation here would be to replace E131-S134 (*i.e.* sequence ETLS) in TMADH with A131-T140 (*i.e.* sequence ATVPGCPGFT) which are the equivalent residues in DMADH. In conjunction with this, the Y60Q mutation is also proposed since it was found to be a key residue with respect to substrate specifity by Basran *et al.* (1997). Therefore, the proposed TMADH 'insertion mutation' is: $Y_{60}Q$ + $_{131}ETLS_{134}$-ATVPGCPGFT. However, there are no current plans for the collaborators to study the effect of this mutation on substrate specificity.

7.4 Conclusion

An attempt was made to automatically determine the electron transfer pathway within TMADH, from the FMN to ETF. The results suggested Glu-439 as a possible residue involved in the pathway. However, the program used for this purpose seems to treat electronic orbitals too simplistically, especially with respect to the 4Fe-4S cluster, to place any reasonable weight on the results. Manual inspection suggests that His-443 could be involved in the electron transfer pathway. A point to note is that Wilson *et al.* (1997), also by manual inspection, suggest that Val-344 could be involved as an alternative to Glu-439.

The structural examination of TMADH with respect to its 17 C-terminal residues indicated that these residues were probably not involved in substrate access, but were most likely to to contribute to the stability of the dimer. However, Ertughrul *et al.* (in press) found that deletion of these residues does not significantly affect either dimer stability or the overall structural integrity of the enzyme. They also report that the deletion of these residues (even that of the 5 C-terminal residues) strongly diminishes the ability of the enzyme to covalently bind the FMN prosthetic group (which is positioned more than 20Å away from the C-terminus). It is thought that the minor conformational changes occurring from the absence of the C-terminal residues in the mutants have an indirect effect on the structure within the substrate binding site, causing a reduction in the ability of the enzyme to covalently bind FMN during the folding of TMADH.

The existence of a central channel along the 2-fold symmetry axis of the TMADH dimer was already known; two additional, larger, symmetric, non-central channels were identified in this study. Field line calculations on the electrostatic potential surface of the channels suggest that TMA is most likely to access the substrate binding site *via* the non-central channels.

In order to understand of TMA/DMA specificity, residues were selected for mutation within TMADH in order to enhance the enzyme's specificity to DMA over TMA, based on (1) structural rationale relating to a hydrogen bonding network in the substrate binding site, and (2) the insertion/deletion region between TMADH and DMADH (as well as the Y60Q aromatic bowl mutation).

The mutations based on the hydrogen bonding network involved three combinations of the following residues: Y60Q, H71L, L73I, S74T, W105F, E131A and T257N. These mutations caused structural perturbations whereby the FMN was not properly incorporated into the substrate binding site. Thus, the role of the hydrogen bonding network residues were not fully rationalised. The proposed mutation based primarily on the insertion/deletion region was as follows: $Y_{60}Q$ + $_{131}ETLS_{134}$-ATVPGCPGFT. However, there are no immediate plans for the collaborators to study the effect on substrate specificity by this mutation.

CHAPTER 8

OVERALL CONCLUSION

Currently, predicted structures are generally assessed not by accuracy but by precision. Values of particular structural features may cluster well with observed values from the available data set and may thus be precise; however, accuracy can only be properly determined if and when the crystal/NMR structure of the protein in question becomes solved.

The RMSDs between models presented in this thesis and the respective crystal structures are plotted against percentage sequence identity in Figure 8.1. The results reflect the fact that prediction-based models do not meet the accuracy of experimentally derived structures unless the percentage sequence identity between the query and template proteins is very high (*e.g.* 85%). Therefore, a larger number of 3D protein structures need to be determined, thereby yielding a more extensive structural data set to draw from, in order for the prediction results to be improved.

However, Figure 8.1 indicates that the models presented in this thesis are comparable in accuracy to those of other modelling groups in the scientific community. Regarding the two entries submitted for the CASP2 comparative modelling category (proteins T0017 and T0024): both of these lie below the average Cα RMSD for their respective participant entries, and the accuracy of these models may therefore be considered to be better than average (the T0017 model approaches close to the minimum RMSD for this target protein).

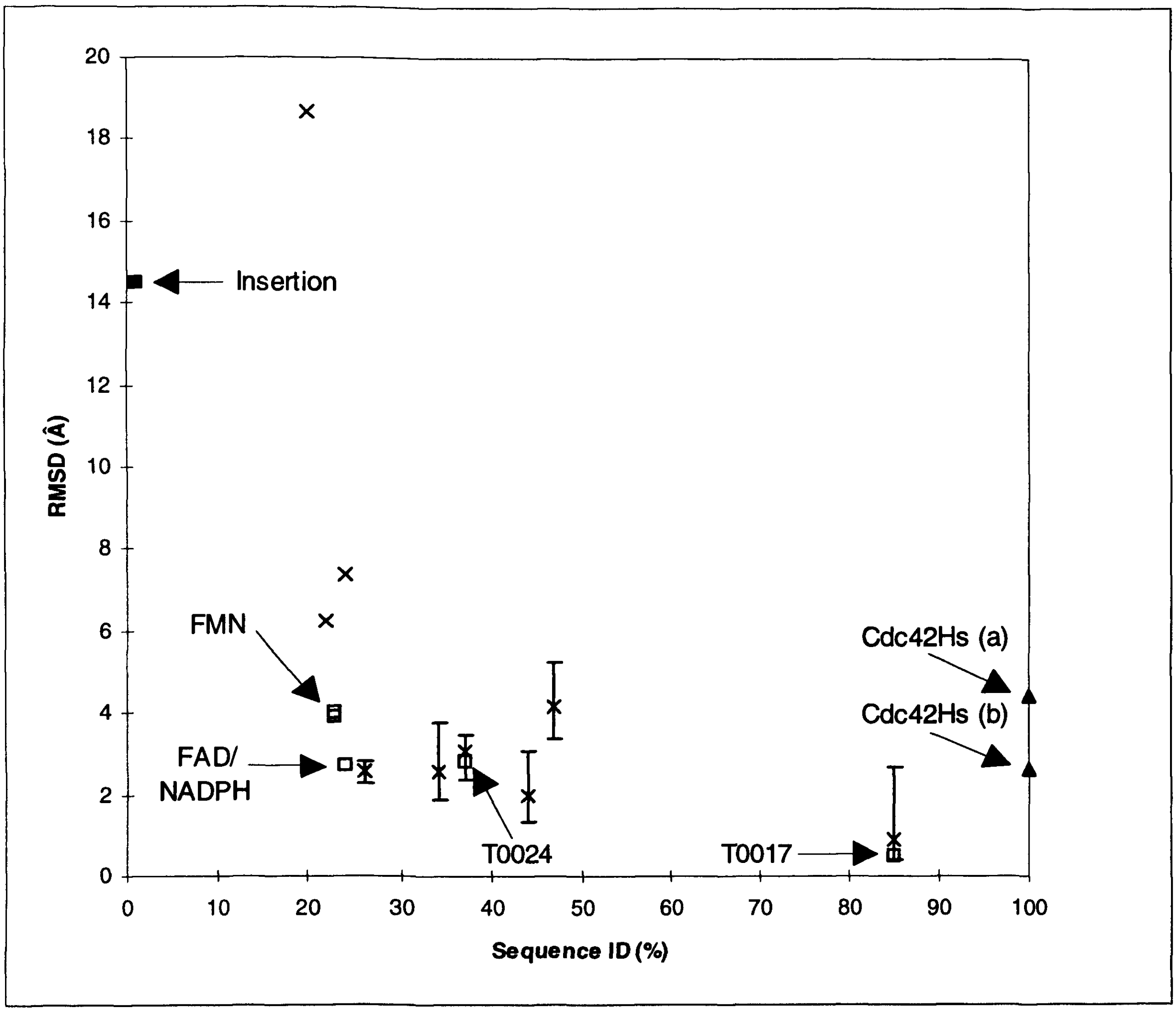


Figure 8.1 Plot of Cα RMSD against percentage sequence identity between query and principal template proteins. × symbols indicate mean values for CASP2 comparative modelling category proteins (Martin *et al.*, 1997) for all groups attempting to model the relevant protein (vertical bars indicate range of values if more than one group attempted to model that protein); this is used as a reference for the accuracy of the models presented in this thesis. □ symbols indicate homology models presented in this thesis (FMN refers to the FMN domain, FAD/NADPH refers to the FAD+FAD/NADPH domain, and T0017 and T0024 are CASP2 codes for proteins modelled by the Sutcliffe group). ■ indicates the threading-based model for the insertion domain of P450R. ▲ symbols indicate the Cα RMSD between the NMR derived structure of Cdc42Hs·GDP and Cdc42Hs·GMPPNP in complex with p50RhoGAP (PDB code 1AM4; Rittinger *et al.*, 1997a), and was included within the graph for the sake of completion (Cdc42Hs(a) refers to Cα RMSD over all common atoms, Cdc42Hs(b) refers to Cα RMSD excluding Switches I and II which were thought to undergo conformational change between GDP- and GTP-bound forms).

The models produced for the FMN and FAD+FAD/NADPH domains of P450R are also as good as would be expected for the relatively low percentage identity with the principal parent protein. There is a rapid increase in Cα RMSD as the percentage identity drops below approximately 25% in Figure 8.1. Thus, the apparently poor Cα RMSD for the insertion domain of P450R is not unexpected.

Although Cdc42Hs·GDP was determined from NMR data rather than from predictive modelling, its structure is included in Figure 8.1 for the sake of completeness. The high Cα RMSD, even when the Switch I and II regions are excluded, could arise from a combination of factors: the experimentally derived structure with which it could most directly be compared is the crystal structure of Cdc42Hs that is (1) GMPPNP-bound and (2) in complex with p50RhoGAP, both of which could result in conformational differences with respect to Cdc42Hs·GDP; the relatively small number of distance constraints available from the NMR data for the Cdc42Hs·GDP, leading to the calculation of the structure in absence of the GDP, could also contribute to the deviation.

For homology-based methods, which depend on generating accurate sequence alignments between query and template proteins, sequence alignment accuracy - the most critical component of producing accurate models - becomes more problematic as sequence identity decreases. However, alignment can be improved in the future if structural as well as sequence information is incorporated into scoring functions.

A threading-based approach is not as reliable as a homology-based approach due to two factors: firstly, determination of the correct fold is less accurate, and secondly, achieving correct sequence alignment is more problematic. However, threading can play a potentially useful role in structure prediction especially when a suitable structural template cannot be found based on sequence similarity. More confidence will be placed on threading if more structural information is taken into account in order to produce improved scoring functions. Note that as more 3D structures are determined, a greater number of proteins (with unknown structure) can be modelled based on sequence homology, and threading will therefore probably become redundant.

For protein-protein interactions, automated docking is unreliable. The main problem for large molecule to large molecule interactions with respect to protein structures is that docking methods have focused on 'ridge-cleft' fitting which is relevant primarily when considering docking into binding pockets, whereas protein-protein interactions tend to occur over relatively large and relatively flat surfaces (Jones & Thronton, 1996), and are hydrophobic rather than hydrophilic in nature in oligomeric structures, for example. However, even within the ridge-cleft fitting category the false positive problem has not been overcome. That is, the lowest energy complexes are not necessarily the correct answer. Therefore, improving inter-molecular scoring functions is a vital area of focus of future work within this field.

Calculations predicting electron transfer pathways serve only as a rough guide to what really takes place, and can sometimes undoubtedly be in error as illustrated by the complete omission of the parameterised 4Fe-4S cluster in the calculated electron pathway within TMADH. Improved parameterisation of chemical groups is therefore an obvious area for future work within this field.

A high resolution experimentally derived 3D protein structure is always preferable to a predicted 3D model. However, in the absence of such an experimentally determined structure, the study of a protein using a 3D structure prediction approach is extremely useful in contributing to the rationale of how a protein functions within its biological context, and forms a powerful basis for subsequent experimental work.

BIBLIOGRAPHY

Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F. and Weng, J. (1987). "Protein Data Bank". *Crystallographic Databases - Information Content, Software Systems, Scientific Applications.* Allen F.H., Bergerhoff, G. and Sievers, R. (eds), Data Commission of the International Union of Crystallography, Cambridge. pp.107-132.

Ahmadian, M.R., Stege, P., Scheffzek, K. and Wittinghofer, A. (1997). "Confirmation of the Arginine Finger Hypothesis for the GAP-stimulated GTP-hydrolysis Reaction of Ras". *Nature Struct. Biol.* 4:686-689.

Altschul, S.F., Gish, W., Miller, W., Myers, W.E. and Lipman, D.J. (1990). "Basic Local Alignment Search Tool". *J. Mol. Biol.* 215:403-410.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997). "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs". *Nucleic Acids Res.* 25:3389-3402.

Attwood, T.K., Beck, M.E., Bleasby, A.J. and Parry-Smith, D.J. (1994). "PRINTS - A Database of Protein Motif Fingerprints". *Nucleic Acids Res.* 22:3590-3596

Bairoch, A., Bucher, P. and Hofmann, K. (1995). "The PROSITE Database, its Status in 1995". *Nucleic Acids Res.* 24:189-196.

Ban, C., Ramakrishnan, B., Ling, K.Y., Kung, C. and Sundaralingham, M. (1994). "Structure of the Recombinant Paramecium-tetraurelia Calmodulin at 1.68 Angstrom Resolution". *Acta Cryst. D* 50:50-63.

Barsukov, I., Modi, S., Lian, L.-Y., Sze, K.H., Paine, M.J.I., Wolf, C.R. and Roberts, G.C.K. (1997). "^{1}H, ^{15}N and ^{13}C NMR Resonance Assignment, Secondary Structure and Global Fold of the FMN-binding Domain of Human Cytochrome P450 Reductase". (1997). *J. Biomol. NMR.* 10:63-65.

Basran, J., Mewies, M., Mathews, F.S. and Scrutton, N.S. (1997). "Selective Modification of Alkylammonium Ion Specificity in Trimethyalamine Dehydrogenase by the Rational Engineering of Cation-π Bonding". *Biochemistry* 36:1989-1998.

Beratan, D.N., Betts, J.N. and Onuchic, J.N. (1991). "Protein Electron Transfer Rates by the Bridging Secondary and Tertiary Structure". *Science* 252:1285-1288.

Bowie, J.U., Luthy, R. and Eisenberg, D. (1991). "A Method to Identify Protein Sequences that Fold into a Known Three-Dimensional Structure". *Science* 253:164-170.

Braun, W. and Go, W. (1985). "Calculation of Protein Conformations by Proton-Proton Distance Constraints: A New Efficient Algorithm". *J. Mol. Biol.* 186:611-626.

Brunger, A.T. (1996). XPLOR software package, version 3.843.

Cabral, J.H.M., Petosa, C., Sutcliffe, M.J., Raza, S., Byron, O., Poy, F., Marfatia, S.M., Chishti, A.H. and Liddington, R.C. (1996). "Crystal Structure of a PDZ Domain". *Nature* 382:649-652.

Chothia, C. (1992). "One Thousand Families for the Molecular Biologist". *Nature* 357:543-544.

Cook, W.J., Jeffrey, L.C., Sullivan, M.L. and Vierstra, R.D. (1992). "Three-Dimensional Structure of a Ubiquitin Conjugating Enzyme (E2)". *J. Biol. Chem.* 267:15116-15121.

Cook, W.J., Jeffrey, L.C., Xu, Y.P. and Chau, V. (1993). "Tertiary Structure of Class I Ubiquitin Conjugating Enzymes are Highly Conserved - Crystal Structure of Yeast UBC4". *Biochemistry* 32:13809-13817.

Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1978). "A Model for Evolutionary Change". *Atlas of Protein Sequence and Structure*. Dayhoff, M.O. (ed). National Biomedical Research Foundation, Washington D.C. 5:345-358.

Doyle, D.A., Lee, A., Lewis, J., Kim, E., Sheng, M. and MacKinnon, R. (1996). "Crystal Structures of a Complexed and Peptide-Free Membrane Protein-Binding Domain: Molecular Basis of Peptide Recognition by PDZ". *Cell* 85:1067-1076.

Dubrochet, J., Adrian, M., Chang, J-J., Homo, J.C., Lepault, J., McDowall, A.W., and Schultz, P. (1988). "Cryo-Electron Microscopy of Vitrified Specimens". *Quart. Rev. Biophys.* 21:129-288.

Ertughrul, O.W.D., Errington, N., Sutcliffe, M.J., Rowe, A.J. and Scrutton, N.S. "Probing the Stabilising Role of C-terminal Residues in Trimethylamine Dehydrogenase". *Protein Eng.* In press.

Feltham, J.L., Dotsch, V., Raza, S., Manor, D., Cerione, R.A., Sutcliffe, M.J., Wagner, G. and Oswald, R.E. (1997). "Definition of the Switch Surface in the Solution Structure of Cdc42Hs". *Biochemistry* 36:8755-8766.

Flaherty, K.M., Zozulya, S., Stryer, L. and McKay, D.B. (1993). "Three-Dimensional Structure of Recoverin, a Calcium Sensor in Vision". *Cell* 75:709-716.

Fukuyama, K., Matsubara, H. and Rogers, L.J. (1992). "Crystal Structure of Oxidised Flavodoxin from a Red Alga Chondrus crispus Refined at 1.8 Angstroms Resolution: Description of the Flavin Mononucleotide Binding Site". *J. Mol. Biol.* 225:775-789.

Garnier, J., Osguthorpe, D.J. and Robson, B. (1978). "Analysis of the Accuracy and Implications of Simple Methods for Predicting the Secondary Structure of Globular Proteins". *J. Mol. Biol.* 120:97-120.

Gilson, K.M., Sharp, K.A. and Honig, B.H. (1988). "Calculating the Electrostatic Potential of Molecules in Solution: Method and Error Assessment". *J. Comp. Chem.* 9:327-335.

Goodford, P.J. (1985). "A Computational Procedure for Determining Energetically Favourable Binding Sites on Biologically Important Macromolecules". *J. Med. Chem.* 28:849-857.

Haggis, G.H. (1966). "The Electron Microscope in Molecular Biology". Longmans. London.

Havel, T.F., Kuntz, I.D. and Crippen, G.M. (1984). "The Theory and Practice of Distance Geometry". *Bull. Math. Biol.* 45:665-720.

Hasemann, C.A., Kurumbail, R.G., Boddupalli, S.S., Peterson, J.A. and Deisenhofer, J. (1995). "Structure and Function of Cytochromes P450: A Comparative Analysis of Three Crystal Structures". *Structure* 3:41-62.

Henikoff, S. and Henikoff, J.G. (1992). "Amino Acid Substitution Matrices from Protein Blocks". *Proc. Natl. Acad. Sci. USA* 89:10915-10919.

Higgins, D.G. and Sharp, P.M. (1988). "CLUSTAL: A Package for Performing Multiple Sequence Alignment on a Microcomputer". *Gene* 73:237-244.

Hill, T.L. (1960). An Introduction to Statistical Thermodynamics. Addison-Wesley Publishing Company, Reading, MA, USA.

Hirshberg, M., Stockley, R.W., Dodson, G. and Webb, M.R. (1997). "The Crystal Structure of Human Rac1, a Member of the Rho-family complexed with a GTP Analogue". *Nature Struct. Biol.* 4:147-151.

Holm, L. and Sander, C. (1994). "The FSSP Database of Structurally Aligned Protein Fold Families". *Nucleic Acids Res.* 22:3600-3609.

Ji, X.H., Johnson, W.W., Sesay, M.A., Dickert, L., Prasad, S.M., Ammon, H.L., Armstrong, R.N. and Gilliland, G.G. (1994). "Structure and Function of the Xenobiotic Substrate-binding Site of a Glutathione-S-transferase as revealed by X-ray Crystallographic Analysis of Product Complexes of the Diasteromers of 9-(S-glutathionyl)-10-hydroxy-9,10-dihydrophenanthrene". *Biochemistry* 33:1043-1052.

Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992a). "The Rapid Generation of Mutation Data Matrices from Protein Sequences". *Comp. App. Bio. Sci.* 8:275-282.

Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992b). "A New Approach to Protein Fold Recognition". *Nature* 358:86-89.

Jones, S. and Thornton, J.M. (1996). "Principles of Protein-Protein Interactions". *Proc. Natl. Acad. Sci. USA* 93:13-20.

Kabsch, W. and Sander, C. (1983). "Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features". *Biopolymers* 22:2577-2637.

Karplus, M. (1959). "Contact Electron-Spin Coupling of Nuclear Magnetic Moments". *J. Chem. Phys.* 30:11-15.

Karplus, P.A., Daniels, M.J. and Herriott, J.R. (1991). "Atomic Structure of Ferredoxin-NADP$^+$ Reductase: Prototype for a Structurally Novel Flavoenzyme Family". *Science* 251:60-66.

Kelley, L.A., Gardner, S.P. and Sutcliffe, M.J. (1996). "An Automated Approach for Clustering an Ensemble of NMR-Derived Protein Structures into Conformationally Related Subfamilies". *Protein Eng.* 9:1063-1065.

Kelley, L.A., Gardner, S.P. and Sutcliffe, M.J. (1997). "An Automated Approach for Defining Core Atoms and Domains in an Ensemble of NMR-Derived Protein Structures". *Protein Eng.* 10:737-741.

Kuszewski, J., Gronenborn, A.M. and Clore, G.M. (1995b). "The Impact of Direct Refinement against Proton Chemical Shifts on Protein Structure Determination by NMR". *J. Mag. Res. B* 107:293-297.

Kuszewski, J., Gronenborn, A.M. and Clore, G.M. (1996). "Improving the Quality of NMR and Crystallographic Protein Structures by means of a Conformational Database Potential Derived from Structure Databases". *Protein Sci.* 5:1067-1080.

Kuszewski, J., Qin, J., Gronenborn, A.M. and Clore, G.M. (1995a). "The Impact of Direct Refinement against ^{13}Cα and ^{13}Cβ Chemical Shifts on Protein Structure Determination by NMR". *J. Mag. Res. Series B* 106:92-96.

Laskowski, R.A. (1995). "SURFNET: A Program for Visualising Molecular Surfaces, Cavities, and Intermolecular Interactions". *J. Mol. Graph.* 13:323-330.

Laskowski, R.A., Luscombe, N.M., Swindells, M.B. and Thornton, J.M. (1996b). "Protein Clefts in Molecular Recognition and Function". *Protein Sci.* 5:2483-2452.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993). "PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures". *J. Appl. Cryst.* 26:283-291.

Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996a). "AQUA and PROCHECK-NMR: Programs for Checking the Quality of Protein Structures Solved by NMR". *J. Biomol. NMR* 8:477-486.

Lemer, C.M.-R., Rooman, M.J., and Wodak, S.J. (1995). "Protein Structure Prediction by Threading Methods: Evaluation of Current Techniques". *Proteins* 23:337-355.

Lesk, A.M. and Chothia, C.H. (1986). "The Response of Protein Structures to Amino Acid Sequence Changes". *Phil. Trans. R. Soc. B* 317:345-356.

Lim, L.W., Shamala, N., Mathews, F.S., Steenkamp, D.J., Hamlin, R. and Xuong, N.H. (1986). "3-Dimensional Structure of the Iron-Sulphur Flavoprotein Trimethylamine Dehydrogenase at 2.4 Angstrom Resolution" *J. Biol. Chem.* 261:15140-15146.

Lu, G.G., Campbell, W.H., Schneider, G. and Lindqvist, Y. (1994). "Crystal Structure of the FAD-Containing Fragment of Corn Nitrate Reductase at 2.5 Angstrom Resolution - Relationship to Other Flavoprotein Reductases". *Structure* 2:809-821.

Ludwig, M.L., Drennan, C.L., Hoover, D., Metzger, A.L., Pattridge, K.A. and Weber, C.H. (1992). "Flavodoxin from Anacystis nidulans: Refinement of Two Forms of the Oxidised Protein". To be published.

Luthy, R., Bowie, J.U. and Eisenberg, D. (1992). "Assessment of Protein Models with Three-Dimensional Profiles". *Nature* 356:83-85.

Marcus, R.A. (1956). "On the Theory of Oxidation-Reduction involving Electron Transfer". *J. Chem. Phys.* 24:966-978.

Marchler-Bauer, A., Levitt, M. and Bryant, S.H. (1997). "A Retrospective Analysis of CASP2 Threading Predictions". *Proteins* Supplement1:83-91.

Martin, A.C.R., MacArthur, M.W. and Thornton, J.M. (1997). "Assessment of Comparative Modelling in CASP2". *Proteins* Supplement1:14-28.

Matsumine, A., Ogai, A., Senda, T., Okumura, N., Satoh, K., Baeg, G.-H., Kawahara, T., Kobayashi, S., Okada, M., Toyoshima, K. and Akiyama, T. (1996). "Binding of APC to the Human Homolog of the Drosophila Discs Large Tumor Suppressor Protein". *Science* 272:1020-1023.

McAllister, D., Smith, J.R. and Diserens, N.J. (1985). Computer Modelling in Electrostatics. Research Studies Press, Wiley, New York.

Michie, A.D., Orengo, C.A. and Thornton, J.M. (1996). "Analysis of Domain Structural Class using an Automated Class Assignment Protocol". *J. Mol. Biol.* 262:168-185.

Mosimann, S., Meleshko, R. and James, M.N.G. (1995). “A Critical Assessment of Comparative Molecular Modelling of Tertiary Structures of Proteins”. *Proteins* 23:301-317.

Moult, J., Peterson, J.T., Judson, R. and Fidelis, K. (1995). “A Large-Scale Experiment to Assess Protein Structure Prediction Methods”. *Proteins* 23:ii-iv.

Moult, J., Hubbard, T., Bryant, S.H., Fidelis, K. and Pederson, J.T. (1997). “Critical Assessment of Methods of Protein Structure Prediction (CASP): Round II”. *Proteins* Supplement1:2-6.

Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995). “SCOP: A Structural Classification of Proteins Database for the Investigation of Sequences and Structures”. *J. Mol. Biol.* 247:536-540.

Needleman, S.B. and Wunsch, C.D. (1970). “A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins”. *J. Mol. Biol.* 48:443-453.

Nicholls, A. and Honig, B. (1991). “A Rapid Finite Difference Algorithm, Utilising Successive Over-Relaxation to Solve the Poisson-Boltzmann Equation”. *J. Comp. Chem.* 12:435-445.

Nicholls, A., Sharp, K. and Honig, B. (1991). “Protein Folding and Association: Insights from the Interfacial and Thermodynamic Properties of Hydrocarbons”. *Proteins* 11:281-296.

Nishida, H., Inaka, K., Yamanaka, M., Kaida, S., Kobayashi, K. and Miki, K. (1995). “Crystal STructure of NADH-Cytochrome b5 Reductase from Pig Liver at 2.4 Angstrom Resolution”. *Biochemistry* 34:2763-2767.

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B. and Thornton, J.M. (1997). “A Hierarchic Classification of Protein Domain Structures”. *Structure* 5:1093-1108.

Pearson, W.R. (1990). "Rapid and Sensitive Sequence Comparison with FASTP and FASTA". *Methods Enzymol.* 183:63-98.

Pearson, W.R. and Lipman, D.J. (1988). "Improved Tools for Biological Sequence Comparison". *Proc. Natl. Acad. Sci. USA* 85:2444-2448.

Porter, T.D. (1991). "An Unusual Yet Strongly Conserved Flavoprotein Reductase in Bacteria and Mammals". *Trends Biochem. Sci.* 16:154-158.

Porter, T.D. and Kasper, C.B. (1986). "NADPH-Cytochrome P-450 Oxidoreductase: Flavin Mononucleotide and Flavin Adenine Dinucleotide Evolved from Different Flavoproteins". *Biochem.* 25:1682-1687.

Raghunathan, S., Chandross, R.J., Kretsinger, R.H., Allison, T.J., Penington, C.J. and Rule, G.S. (1994). "Crystal Structure of Human Class Mu Glutathione Transferase GSTM2-2 - Effects of Lattice Packing on Conformational Heterogeneity". *J. Mol. Bol.* 238:815-832.

Rao, S.T., Shaffie, F., Yu, C., Satyshur, K.A., Stockman, B.J. and Sundaralingham, M. (1992). "Structure of the Oxidised Long-Chain Flavaodoxin from Anabaena 7120 at 2 Angstroms Resolution". *Protein Sci.* 1:1413-1427.

Rhodes, G. (1993). "Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models". Academic Press Inc. San Diego.

Rittinger, K., Walker, P.A., Eccleston, J.F., Nurmahomed, K., Owen, D., Laue, E., Gamblin, S.J. and Smerdon, S.J. (1997a). "Crystal Structure of a Small G-protein in Complex with the GTPase-Activating Protein RhoGAP". *Nature* 388:693-697.

Rittinger, K., Walker, P.A., Eccleston, J.F., Smerdon, S.J. and Gamblin, S.J. (1997b). "Structure at 1.65 Angstroms of RhoA and its GTPase-Activating Protein in complex with a Transition State Analogue". *Nature* 389:758-762.

Robson, B. and Suzuki, E. (1976). "Conformational Properties of Amino Acid Residues in Globular Proteins". *J. Mol. Biol.* 107:327-356.

Rost, B. (1995a). "TOPITS: Threading One-dimensional Predictions Into Three-dimensional Structures". *Third International Conference on Intelligent Systems for Molecular Biology, Cambridge, UK*, AAAI Press, CA, USA.

Rost, B. (1995b). "Fitting 1D Predictions into 3D Structures". *Protein Folds: A Distance Based Approach.* Bohr, H. and Brunak, S. (eds), CRC Press, Boca Raton. pp.132-151.

Rost, B. and Sander, C. (1993a). "Improved Prediction of Protein Secondary Structure by Use of Sequence Profiles and Neural Networks". *Proc. Natl. Acad. Sci. USA* 90:7558-7562.

Rost, B. and Sander, C. (1993b). "Prediction of Protein Structure at Better than 70% Accuracy". *J. Mol. Biol.* 232:584-599.

Rost, B. and Sander, C. (1994a). "Combining Evolutionary Information and Neural Networks to Predict Protein Secondary Structure". *Proteins: Struct. Func. Gen.* 19:55-72.

Rost, B. and Sander, C. (1994b). "Conservation and Prediction of Solvent Accessibility in Protein Families". *Proteins* 20:216-226.

Rost, B., Sander, C. and Schneider, R. (1994). "PHD - An Automatic Mail Server for Protein Secondary Structure Prediction". *Comp. App. Bio. Sci.* 10:53-60.

Sali, A. and Blundell, T.L. (1993). "Comparative Protein Modelling by Satisfaction of Spatial Restraints". *J. Mol. Biol.* 234:779-815.

Sali, A., Overington, J.P., Johnson, M.S. and Blundell, T.L. (1990). "From Comparisons of Protein Sequences and Structures to Protein Modelling and Design". *Trends Biochem. Sci.* 15:235-240.

Sander, C. and Schneider, R. (1991). "Database of Homology-Derived Protein Structures and the Structural Meaning of Sequence Alignment". *Proteins* 9:56-68.

Satyshur, K.A., Pyzalska, D., Greaser, M., Rao, S.T. and Sundaralingham, M. (1994). "Structure of Chicken Skeletal-Muscle Troponin-C at 1.78 Angstrom Resolution". *Acta Cryst. D* 50:40-49.

Scheffzek, K., Lautwein, A., Kabsch, W., Ahmadian, M.R. and Wittinghofer, A. (1996). "Crystal Structure of the GTPase-Activating Domain of Human p120GAP and Implications for the Interaction with Ras". *Nature* 384:591-596.

Slichter, C.P. (1993). "Principles of Magnetic Resonance". Springer-Verlag. Berlin.

Shoichet, B.K. and Kuntz, I.D. (1991). "Protein Docking and Complementarity". *J. Mol. Biol.* 221:327-346.

Sippl, M.J. (1990). "Calculation of Conformational Ensembles from Potentials of Mean Force: An Approach to the Knowledge-Based Prediction of Local Structures in Globular Proteins". *J. Mol. Biol.* 213:859-883.

Smart, O., Goodfellow, J.M. and Wallace, B.A. (1993). "The Pore Dimensions of Gramicidin A". *Biophys. J.* 65:2455-2460.

Smith, J.F. and Waterman, M.S. (1981). "Identification of Common Molecular Subsequences". *J. Mol. Biol.* 147:195-197.

Smith, W.W., Burnett, R.M., Darling, G.D. and Ludwig, M.L. (1977). "Structure of the Semiquinone Form of Colstridium MP. Extension of 1.8 Angstroms Resolution and some Comparisons with Oxidised State". *J. Mol. Biol.* 117:195-225.

Spera, S. and Bax, A. (1991). "Empirical Correlation between Protein Backbone Conformation and Cα and Cβ 13C Nuclear Magnetic Resonance Chemical Shifts". *J. Am. Chem. Soc.* 113:5490-5492.

Sutcliffe, M.J., Feltham, J., Cerione, R.A. and Oswald, R.E. (1994). "Model Building Predicts an Additional Conformational Switch when GTP Binds to the Cdc42Hs Protein". *Protein Pep. Lett.* 1:84-91.

Sutcliffe, M.J., Haneef, I., Carney, D. and Blundell, T.L. (1987a). "Knowledge Based Modelling of Homologous Proteins, Part I: Three-Dimensional Frameworks Derived from the Simultaneous Superposition of Multiple Structures". *Protein Eng.* 1:377-384.

Sutcliffe, M.J., Hayes, F.R.F. and Blundell, T.L. (1987b). "Knowledge Based Modelling of Homologous Proteins, Part II: Rules for the Conformations of Substituted Sidechains". *Protein Eng.* 1:385-392.

Taylor, W. (1988). "A Flexible Method to Align Large Numbers of Biological Sequences". *J. Mol. Evol.* 28:161-169.

Tesmer, J.J., Berman, D.M., Gilman, A.G. and Sprang, S.R. (1997). "Structure of RGS4 bound to AlF_4-activated $G_{i\alpha 1}$: Stabilisation of the Transition State for GTP Hydrolysis". *Cell* 89:251-261.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994). "CLUSTALW: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice". *Nucl. Acids Res.* 22:4673-4680.

Tong, L., de Vos, A.M., Milburn, M.V. and Kim, S.-H. (1991). "Crystal Structure at 2.2 Angstroms Resolution of the Catalytic Domains of Normal Ras Protein and an Oncogenic Mutant Complexed with GDP". *J. Mol. Biol.* 217:503-516.

Vermilion, J.L., Ballou, D.P., Massey, V. and Coon, M.J. (1981). "Role of Flavins in NADPH-Cytochrome P-450 Reductase". *J. Biol. Chem.* 256:266-277.

Wang, Z-X. (1996). "How Many Fold Types of Protein Are There in Nature?". *Proteins* 26:186-191.

Wang, M., Roberts, D.L., Paschke, R., Shea, T.M., Masters, B.S.S. and Kim, J.-J.P. (1997). "Three-Dimensional Structure of NADPH-Cytochrome P450 Reductase: Prototype for FMN- and FAD-containing Enzymes". *Proc. Natl. Acad. Sci. USA* 94:8411-8416.

Watt, W., Tulinsky, A., Swenson, R.P. and Watenpaugh, K.D. (1991). "Comparison of the Crystal Structures of a Flavodoxin in its Three Oxidation States at Cryogenic Temperatures". *J. Mol. Biol.* 218:195-208.

Wilmot, C.M. and Thornton, J.M. (1990). "β-Turns and their Distortions: A Proposed New Nomenclature". *Protein Eng.* 3:479-493.

Wilson, E.K., Huang, L., Sutcliffe, M.J., Mathews, F.S., Hille, R. and Scrutton, N.S. (1997). "An Exposed Tyrosine on the Surface of Trimethylamine Dehydrogenase Facilitates Electron Transfer to Electron Transferring Flavoprotein: Kinetics of Transfer in Wild-Type and Mutant Complexes". *Biochemistry* 36:41-48.

Wilson, E.K., Mathews, F.S., Packman, L.C. and Scrutton, N.S. (1995). "Electron Tunnelling in Substrate-Reduced Trimethylamine Dehydrogenase: Kinetics of Electron Transfer and Analysis of Tunnelling Pathway". *Biochemistry* 34:2584-2591.

Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992). "The Chemical Shift Index – A Fast and Simple Method for the Assignment of Protein Secondary Structure through NMR Spectroscopy". *Biochemistry* 31:1647-1651.

Wishart, D.S. and Sykes, B.D. (1994). "The ^{13}C Chemical Shift Index: A Simple Method for the Identification of Protein Secondary Structure using ^{13}C Chemical Shift Data". *J. Biomol. NMR* 4:171-180.

Wu, W.J., Leonard, D.A., Paar, J.M., Cerione, R. and Manor, D. (1997). "Interaction between Cdc42Hs and RhoGDI is mediated through the Rho Insert Region". *J. Biol. Chem.* 272:26153-26158.

Zhao, Q., Modi, S., Smith, G., Paine, M., McDonagh, P.D., Wolf, C.R., Tew, D., Lian, L.-Y., Roberts, G.C.K. and Driessen, H.P.C. (1996). "Crystallisation and Preliminary X-ray Diffraction Studies of Human Cytochrome-P450 Reductase". *J. Struct. Biol.* 116:320-325.

APPENDIX

Primary NMR Data for Cdc42Hs

Distance Restraints

```
!INTRARESIDUE AND SEQUENTIAL HN OR HA TO HN OR HA

assi (resi  -5 and name HN ) (resi  -5 and name HA ) 3.4 3.4 0.0
assi (resi  -5 and name HN ) (resi  -6 and name HA ) 2.4 2.4 0.0
assi (resi  -4 and name HN ) (resi  -5 and name HA ) 2.4 2.4 0.0
assi (resi  -4 and name HN ) (resi  -4 and name HA ) 2.9 2.9 0.0
assi (resi  -3 and name HN ) (resi  -4 and name HA ) 2.4 2.4 0.0
assi (resi  -3 and name HN ) (resi  -3 and name HD# ) 2.4 2.4 0.0
assi (resi  -3 and name HA ) (resi  -3 and name HB ) 3.4 3.4 0.0
assi (resi  -3 and name HA ) (resi  -3 and name HG1# ) 5.5 5.5 0.6
assi (resi  -3 and name HA ) (resi  -3 and name HG1# ) 2.9 2.9 0.6
assi (resi  -3 and name HA ) (resi  -3 and name HG2# ) 2.9 2.9 1.1
assi (resi  -3 and name HA ) (resi  -3 and name HD# ) 2.9 2.9 0.0
assi (resi  -2 and name HN ) (resi  -3 and name HA ) 2.4 2.4 0.0
assi (resi  -1 and name HN ) (resi  -2 and name HA ) 2.4 2.4 0.0
assi (resi   1 and name HN ) (resi  -1 and name HA ) 2.4 2.4 0.0
assi (resi   2 and name HN ) (resi   2 and name HA ) 2.9 2.9 0.0
assi (resi   4 and name HN ) (resi   5 and name HN ) 4.0 4.0 0.0
assi (resi   4 and name HN ) (resi   4 and name HA ) 4.0 4.0 0.0
assi (resi   5 and name HN ) (resi   4 and name HA ) 2.9 2.9 0.0
assi (resi   6 and name HN ) (resi   6 and name HA ) 4.0 4.0 0.0
assi (resi   6 and name HN ) (resi   5 and name HA ) 2.9 2.9 0.0
assi (resi   7 and name HN ) (resi   6 and name HN ) 5.5 5.5 0.0
assi (resi   7 and name HN ) (resi   8 and name HN ) 5.5 5.5 0.0
assi (resi   7 and name HN ) (resi   6 and name HA ) 2.9 2.9 0.0
assi (resi   7 and name HN ) (resi   7 and name HA ) 4.0 4.0 0.0
assi (resi   8 and name HN ) (resi   9 and name HN ) 5.5 5.5 0.0
assi (resi   8 and name HN ) (resi   7 and name HA ) 2.9 2.9 0.0
assi (resi   8 and name HN ) (resi   8 and name HA ) 2.9 2.9 0.0
assi (resi   9 and name HN ) (resi   8 and name HA ) 2.4 2.4 0.0
assi (resi  10 and name HN ) (resi   9 and name HN ) 5.5 5.5 0.0
assi (resi  10 and name HN ) (resi   9 and name HA ) 3.4 3.4 0.0
assi (resi  12 and name HN ) (resi  13 and name HN ) 4.0 4.0 0.0
assi (resi  12 and name HN ) (resi  12 and name HA# ) 2.9 2.9 1.0
assi (resi  12 and name HN ) (resi  12 and name HA# ) 2.9 2.9 1.0
assi (resi  13 and name HN ) (resi  14 and name HN ) 3.4 3.4 0.0
assi (resi  13 and name HN ) (resi  13 and name HA ) 2.9 2.9 0.0
assi (resi  13 and name HN ) (resi  12 and name HA# ) 2.9 2.9 1.0
assi (resi  13 and name HN ) (resi  12 and name HA# ) 2.9 2.9 1.0
assi (resi  15 and name HN ) (resi  16 and name HN ) 3.4 3.4 0.0
assi (resi  15 and name HN ) (resi  14 and name HN ) 5.5 5.5 0.0
assi (resi  15 and name HN ) (resi  12 and name HA# ) 4.0 4.0 1.0
assi (resi  16 and name HN ) (resi  16 and name HA ) 4.0 4.0 0.0
assi (resi  17 and name HN ) (resi  16 and name HN ) 4.0 4.0 0.0
assi (resi  17 and name HN ) (resi  18 and name HN ) 5.5 5.5 0.0
assi (resi  17 and name HN ) (resi  17 and name HA ) 4.0 4.0 0.0
assi (resi  17 and name HN ) (resi  16 and name HA ) 4.0 4.0 0.0
assi (resi  18 and name HN ) (resi  19 and name HN ) 3.4 3.4 0.0
assi (resi  19 and name HN ) (resi  19 and name HA ) 2.9 2.9 0.0
assi (resi  19 and name HN ) (resi  18 and name HA ) 3.4 3.4 0.0
assi (resi  20 and name HN ) (resi  21 and name HN ) 3.4 3.4 0.0
assi (resi  20 and name HN ) (resi  19 and name HN ) 3.4 3.4 0.0
assi (resi  20 and name HN ) (resi  20 and name HA ) 2.9 2.9 0.0
assi (resi  21 and name HN ) (resi  22 and name HN ) 3.4 3.4 0.0
assi (resi  21 and name HN ) (resi  21 and name HA ) 3.4 3.4 0.0
assi (resi  22 and name HN ) (resi  23 and name HN ) 3.4 3.4 0.0
assi (resi  22 and name HN ) (resi  21 and name HA ) 3.4 3.4 0.0
assi (resi  22 and name HA ) (resi  22 and name HN ) 3.4 3.4 0.0
assi (resi  23 and name HN ) (resi  24 and name HN ) 3.4 3.4 0.0
assi (resi  23 and name HN ) (resi  23 and name HA ) 2.9 2.9 0.0
assi (resi  24 and name HN ) (resi  25 and name HN ) 2.9 2.9 0.0
assi (resi  24 and name HN ) (resi  24 and name HA ) 2.9 2.9 0.0
assi (resi  25 and name HN ) (resi  25 and name HA ) 2.9 2.9 0.0
assi (resi  25 and name HN ) (resi  24 and name HA ) 3.4 3.4 0.0
assi (resi  26 and name HN ) (resi  25 and name HN ) 2.9 2.9 0.0
assi (resi  26 and name HN ) (resi  26 and name HA ) 2.4 2.4 0.0
assi (resi  27 and name HN ) (resi  27 and name HA ) 2.9 2.9 0.0
assi (resi  28 and name HN ) (resi  27 and name HN ) 3.4 3.4 0.0
assi (resi  28 and name HN ) (resi  28 and name HA ) 2.9 2.9 0.0
assi (resi  28 and name HN ) (resi  27 and name HA ) 2.4 2.4 0.0
assi (resi  31 and name HN ) (resi  31 and name HA ) 2.9 2.9 0.0
assi (resi  32 and name HN ) (resi  31 and name HN ) 2.9 2.9 0.0
assi (resi  32 and name HN ) (resi  31 and name HA ) 2.4 2.4 0.0
assi (resi  33 and name HN ) (resi  32 and name HA ) 2.4 2.4 0.0
assi (resi  33 and name HN ) (resi  33 and name HA ) 3.4 3.4 0.0
assi (resi  42 and name HN ) (resi  41 and name HA ) 2.4 2.4 0.0
assi (resi  42 and name HN ) (resi  42 and name HA ) 2.9 2.9 0.0
assi (resi  43 and name HN ) (resi  44 and name HN ) 4.0 4.0 0.0
assi (resi  43 and name HN ) (resi  43 and name HA ) 2.9 2.9 0.0
assi (resi  43 and name HN ) (resi  42 and name HA ) 2.4 2.4 0.0
assi (resi  44 and name HN ) (resi  43 and name HA ) 2.4 2.4 0.0
assi (resi  44 and name HN ) (resi  44 and name HA ) 3.4 3.4 0.0
assi (resi  45 and name HN ) (resi  44 and name HN ) 4.0 4.0 0.0
assi (resi  45 and name HN ) (resi  46 and name HN ) 4.0 4.0 0.0
assi (resi  45 and name HN ) (resi  45 and name HA ) 2.9 2.9 0.0
assi (resi  45 and name HN ) (resi  44 and name HA ) 2.4 2.4 0.0
assi (resi  46 and name HN ) (resi  45 and name HA ) 2.4 2.4 0.0
assi (resi  46 and name HN ) (resi  46 and name HA ) 2.9 2.9 0.0
assi (resi  47 and name HN ) (resi  48 and name HN ) 2.9 2.9 0.0
assi (resi  47 and name HN ) (resi  47 and name HA# ) 2.4 2.4 1.0
assi (resi  47 and name HN ) (resi  47 and name HA# ) 2.4 2.4 1.0
assi (resi  48 and name HN ) (resi  49 and name HN ) 2.9 2.9 0.0
assi (resi  48 and name HN ) (resi  47 and name HA# ) 2.9 2.9 1.0
assi (resi  48 and name HN ) (resi  48 and name HA# ) 2.4 2.4 1.0
assi (resi  48 and name HN ) (resi  47 and name HA# ) 2.4 2.4 1.0
assi (resi  48 and name HN ) (resi  48 and name HA# ) 2.4 2.4 1.0
assi (resi  48 and name HA# ) (resi  48 and name HA# ) 2.9 2.9 1.0
assi (resi  49 and name HN ) (resi  49 and name HA ) 2.9 2.9 0.0
assi (resi  49 and name HN ) (resi  48 and name HA# ) 3.4 3.4 1.0
assi (resi  49 and name HN ) (resi  48 and name HA# ) 2.4 2.4 1.0
assi (resi  49 and name HA ) (resi  48 and name HA# ) 3.4 3.4 1.0
assi (resi  51 and name HN ) (resi  50 and name HA ) 2.4 2.4 0.0
assi (resi  51 and name HN ) (resi  51 and name HA ) 3.4 3.4 0.0
assi (resi  52 and name HN ) (resi  53 and name HN ) 5.5 5.5 0.0
assi (resi  52 and name HN ) (resi  51 and name HA ) 2.4 2.4 0.0
assi (resi  53 and name HN ) (resi  53 and name HA ) 2.9 2.9 0.0
assi (resi  53 and name HN ) (resi  52 and name HA ) 2.4 2.4 0.0
assi (resi  54 and name HN ) (resi  55 and name HN ) 4.0 4.0 0.0
assi (resi  54 and name HN ) (resi  53 and name HA ) 2.4 2.4 0.0
assi (resi  54 and name HN ) (resi  54 and name HA# ) 2.9 2.9 1.0
assi (resi  54 and name HN ) (resi  54 and name HA# ) 2.9 2.9 1.0
assi (resi  55 and name HN ) (resi  54 and name HA# ) 2.9 2.9 1.0
assi (resi  55 and name HN ) (resi  54 and name HA# ) 3.4 3.4 1.0
assi (resi  56 and name HN ) (resi  55 and name HA ) 2.9 2.9 0.0
assi (resi  56 and name HN ) (resi  56 and name HA ) 4.0 4.0 0.0
assi (resi  57 and name HN ) (resi  56 and name HA ) 2.9 2.9 0.0
assi (resi  57 and name HN ) (resi  57 and name HA ) 4.0 4.0 0.0
assi (resi  61 and name HN ) (resi  62 and name HN ) 4.0 4.0 0.0
assi (resi  61 and name HN ) (resi  61 and name HA ) 2.9 2.9 0.0
assi (resi  62 and name HN ) (resi  62 and name HA ) 2.4 2.4 0.0
assi (resi  63 and name HN ) (resi  62 and name HN ) 3.4 3.4 0.0
assi (resi  63 and name HN ) (resi  64 and name HN ) 2.9 2.9 0.0
assi (resi  63 and name HN ) (resi  63 and name HA ) 2.4 2.4 0.0
assi (resi  64 and name HN ) (resi  64 and name HA ) 2.4 2.4 0.0
assi (resi  70 and name HN ) (resi  71 and name HN ) 4.0 4.0 0.0
assi (resi  71 and name HN ) (resi  71 and name HA ) 2.9 2.9 0.0
assi (resi  72 and name HN ) (resi  71 and name HN ) 4.0 4.0 0.0
assi (resi  74 and name HN ) (resi  75 and name HN ) 4.0 4.0 0.0
assi (resi  74 and name HN ) (resi  74 and name HA ) 2.9 2.9 0.0
assi (resi  75 and name HN ) (resi  74 and name HA ) 2.9 2.9 0.0
assi (resi  77 and name HN ) (resi  76 and name HN ) 2.9 2.9 0.0
assi (resi  77 and name HN ) (resi  78 and name HN ) 4.0 4.0 0.0
assi (resi  77 and name HN ) (resi  76 and name HA ) 2.9 2.9 0.0
assi (resi  77 and name HN ) (resi  77 and name HA ) 3.4 3.4 0.0
assi (resi  79 and name HN ) (resi  78 and name HN ) 5.5 5.5 0.0
assi (resi  79 and name HN ) (resi  78 and name HA ) 2.4 2.4 0.0
assi (resi  80 and name HN ) (resi  79 and name HA ) 2.4 2.4 0.0
assi (resi  81 and name HN ) (resi  81 and name HA ) 3.4 3.4 0.0
assi (resi  81 and name HN ) (resi  80 and name HA ) 2.4 2.4 0.0
assi (resi  82 and name HN ) (resi  81 and name HA ) 2.4 2.4 0.0
assi (resi  83 and name HN ) (resi  83 and name HA ) 2.9 2.9 0.0
assi (resi  84 and name HN ) (resi  85 and name HN ) 3.4 3.4 0.0
assi (resi  84 and name HN ) (resi  83 and name HA ) 2.9 2.9 0.0
assi (resi  84 and name HN ) (resi  84 and name HA ) 2.9 2.9 0.0
assi (resi  85 and name HN ) (resi  85 and name HA ) 2.9 2.9 0.0
assi (resi  86 and name HN ) (resi  86 and name HA ) 2.9 2.9 0.0
assi (resi  86 and name HN ) (resi  85 and name HA ) 3.4 3.4 0.0
assi (resi  88 and name HN ) (resi  88 and name HA ) 2.4 2.4 0.0
assi (resi  89 and name HN ) (resi  88 and name HN ) 4.0 4.0 0.0
assi (resi  89 and name HN ) (resi  89 and name HA ) 4.0 4.0 0.0
assi (resi  89 and name HN ) (resi  88 and name HA ) 2.9 2.9 0.0
assi (resi  90 and name HN ) (resi  91 and name HN ) 2.9 2.9 0.0
assi (resi  90 and name HN ) (resi  89 and name HN ) 2.9 2.9 0.0
assi (resi  90 and name HN ) (resi  90 and name HA ) 2.9 2.9 0.0
assi (resi  91 and name HN ) (resi  92 and name HN ) 2.9 2.9 0.0
assi (resi  91 and name HN ) (resi  91 and name HA ) 3.4 3.4 0.0
assi (resi  92 and name HN ) (resi  92 and name HA ) 2.9 2.9 0.0
assi (resi  93 and name HN ) (resi  94 and name HN ) 2.9 2.9 0.0
assi (resi  93 and name HN ) (resi  92 and name HA ) 3.4 3.4 0.0
assi (resi  93 and name HN ) (resi  93 and name HA ) 2.9 2.9 0.0
assi (resi  94 and name HN ) (resi  94 and name HA ) 2.9 2.9 0.0
assi (resi  95 and name HN ) (resi  96 and name HN ) 2.9 2.9 0.0
assi (resi  95 and name HN ) (resi  94 and name HN ) 2.9 2.9 0.0
assi (resi  95 and name HN ) (resi  95 and name HA ) 2.9 2.9 0.0
assi (resi  95 and name HN ) (resi  94 and name HA ) 3.4 3.4 0.0
assi (resi  96 and name HN ) (resi  96 and name HA ) 2.9 2.9 0.0
assi (resi  97 and name HN ) (resi  98 and name HN ) 3.4 3.4 0.0
assi (resi  97 and name HN ) (resi  96 and name HN ) 2.9 2.9 0.0
assi (resi  97 and name HN ) (resi  96 and name HA ) 3.4 3.4 0.0
assi (resi 100 and name HN ) (resi 100 and name HA ) 2.9 2.9 0.0
assi (resi 101 and name HN ) (resi 100 and name HN ) 3.4 3.4 0.0
assi (resi 101 and name HN ) (resi 101 and name HA ) 2.9 2.9 0.0
assi (resi 102 and name HN ) (resi 101 and name HN ) 4.0 4.0 0.0
assi (resi 102 and name HN ) (resi 102 and name HA ) 2.9 2.9 0.0
assi (resi 103 and name HN ) (resi 104 and name HN ) 2.9 2.9 0.0
assi (resi 103 and name HN ) (resi 103 and name HA ) 2.9 2.9 0.0
assi (resi 104 and name HN ) (resi 104 and name HA ) 2.9 2.9 0.0
assi (resi 105 and name HN ) (resi 104 and name HN ) 2.9 2.9 0.0
assi (resi 105 and name HN ) (resi 105 and name HA ) 2.9 2.9 0.0
assi (resi 105 and name HN ) (resi 104 and name HA ) 3.4 3.4 0.0
assi (resi 107 and name HN ) (resi 108 and name HN ) 2.4 2.4 0.0
assi (resi 107 and name HN ) (resi 107 and name HA ) 2.4 2.4 0.0
assi (resi 108 and name HN ) (resi 108 and name HA ) 2.9 2.9 0.0
assi (resi 108 and name HN ) (resi 107 and name HA ) 3.4 3.4 0.0
assi (resi 110 and name HN ) (resi 110 and name HA ) 2.9 2.9 0.0
assi (resi 110 and name HN ) (resi 109 and name HA ) 2.4 2.4 0.0
assi (resi 111 and name HN ) (resi 110 and name HA ) 2.9 2.9 0.0
assi (resi 111 and name HN ) (resi 111 and name HA ) 2.9 2.9 0.0
assi (resi 112 and name HN ) (resi 113 and name HN ) 4.0 4.0 0.0
assi (resi 112 and name HN ) (resi 112 and name HA ) 3.4 3.4 0.0
assi (resi 112 and name HN ) (resi 111 and name HA ) 2.4 2.4 0.0
assi (resi 113 and name HN ) (resi 112 and name HA ) 2.4 2.4 0.0
assi (resi 114 and name HN ) (resi 113 and name HA ) 2.9 2.9 0.0
assi (resi 114 and name HN ) (resi 114 and name HA# ) 4.0 4.0 1.0
assi (resi 115 and name HN ) (resi 114 and name HA# ) 2.9 2.9 1.0
assi (resi 116 and name HN ) (resi 116 and name HA ) 3.4 3.4 0.0
assi (resi 117 and name HN ) (resi 116 and name HN ) 3.4 3.4 0.0
assi (resi 117 and name HN ) (resi 118 and name HN ) 3.4 3.4 0.0
assi (resi 118 and name HN ) (resi 119 and name HN ) 2.9 2.9 0.0
assi (resi 118 and name HN ) (resi 117 and name HA ) 2.9 2.9 0.0
assi (resi 119 and name HN ) (resi 120 and name HN ) 2.9 2.9 0.0
assi (resi 119 and name HN ) (resi 119 and name HA ) 2.4 2.4 0.0
assi (resi 120 and name HN ) (resi 119 and name HA ) 3.4 3.4 0.0
assi (resi 120 and name HN ) (resi 120 and name HA ) 2.9 2.9 0.0
assi (resi 121 and name HN ) (resi 121 and name HA ) 2.9 2.9 0.0
assi (resi 121 and name HN ) (resi 120 and name HA ) 3.4 3.4 0.0
assi (resi 122 and name HN ) (resi 122 and name HA ) 2.9 2.9 0.0
assi (resi 122 and name HN ) (resi 121 and name HA ) 2.9 2.9 0.0
assi (resi 124 and name HN ) (resi 125 and name HN ) 2.9 2.9 0.0
assi (resi 124 and name HN ) (resi 124 and name HA ) 2.4 2.4 0.0
assi (resi 125 and name HN ) (resi 124 and name HA ) 3.4 3.4 0.0
assi (resi 126 and name HN ) (resi 125 and name HN ) 4.0 4.0 0.0
assi (resi 126 and name HN ) (resi 126 and name HA ) 2.9 2.9 0.0
assi (resi 126 and name HB ) (resi 126 and name HA ) 4.0 4.0 0.0
assi (resi 127 and name HN ) (resi 128 and name HN ) 2.9 2.9 0.0
assi (resi 127 and name HN ) (resi 126 and name HA ) 3.4 3.4 0.0
assi (resi 128 and name HN ) (resi 128 and name HA ) 2.4 2.4 0.0
assi (resi 129 and name HN ) (resi 130 and name HN ) 2.9 2.9 0.0
assi (resi 129 and name HN ) (resi 128 and name HN ) 2.9 2.9 0.0
assi (resi 129 and name HN ) (resi 129 and name HA ) 2.9 2.9 0.0
assi (resi 130 and name HN ) (resi 129 and name HA ) 3.4 3.4 0.0
assi (resi 130 and name HN ) (resi 130 and name HA ) 2.4 2.4 0.0
assi (resi 131 and name HN ) (resi 130 and name HN ) 2.9 2.9 0.0
assi (resi 131 and name HN ) (resi 132 and name HN ) 2.4 2.4 0.0
assi (resi 131 and name HN ) (resi 132 and name HA ) 5.5 5.5 0.0
assi (resi 131 and name HN ) (resi 131 and name HA ) 2.4 2.4 0.0
assi (resi 132 and name HN ) (resi 133 and name HN ) 2.4 2.4 0.0
assi (resi 132 and name HN ) (resi 132 and name HA ) 2.4 2.4 0.0
assi (resi 132 and name HN ) (resi 131 and name HA ) 2.9 2.9 0.0
assi (resi 133 and name HN ) (resi 132 and name HA ) 2.9 2.9 0.0
assi (resi 133 and name HN ) (resi 133 and name HA ) 2.4 2.4 0.0
assi (resi 134 and name HN ) (resi 134 and name HA ) 2.9 2.9 0.0
assi (resi 134 and name HN ) (resi 133 and name HA ) 2.9 2.9 0.0
assi (resi 135 and name HN ) (resi 135 and name HA ) 2.4 2.4 0.0
assi (resi 137 and name HN ) (resi 136 and name HA ) 2.4 2.4 0.0
assi (resi 137 and name HN ) (resi 137 and name HA ) 2.9 2.9 0.0
assi (resi 138 and name HN ) (resi 137 and name HN ) 5.5 5.5 0.0
assi (resi 138 and name HN ) (resi 137 and name HA ) 2.4 2.4 0.0
assi (resi 140 and name HN ) (resi 141 and name HN ) 2.9 2.9 0.0
assi (resi 140 and name HN ) (resi 140 and name HA ) 2.9 2.9 0.0
assi (resi 141 and name HN ) (resi 140 and name HA ) 2.9 2.9 0.0
assi (resi 141 and name HN ) (resi 141 and name HA ) 2.9 2.9 0.0
assi (resi 142 and name HN ) (resi 143 and name HN ) 2.9 2.9 0.0
assi (resi 142 and name HN ) (resi 142 and name HA ) 2.9 2.9 0.0
assi (resi 142 and name HN ) (resi 141 and name HA ) 3.4 3.4 0.0
assi (resi 143 and name HN ) (resi 142 and name HA ) 3.4 3.4 0.0
assi (resi 143 and name HN ) (resi 143 and name HA ) 2.9 2.9 0.0
assi (resi 144 and name HN ) (resi 143 and name HN ) 2.9 2.9 0.0
assi (resi 144 and name HN ) (resi 144 and name HA ) 2.9 2.9 0.0
assi (resi 145 and name HN ) (resi 146 and name HN ) 2.4 2.4 0.0
assi (resi 145 and name HN ) (resi 145 and name HA ) 2.9 2.9 0.0
assi (resi 146 and name HN ) (resi 147 and name HN ) 2.9 2.9 0.0
assi (resi 146 and name HN ) (resi 145 and name HA ) 3.4 3.4 0.0
assi (resi 146 and name HN ) (resi 146 and name HA ) 2.9 2.9 0.0
assi (resi 147 and name HN ) (resi 148 and name HN ) 2.4 2.4 0.0
assi (resi 147 and name HN ) (resi 147 and name HA ) 2.4 2.4 0.0
assi (resi 147 and name HN ) (resi 146 and name HA ) 2.9 2.9 0.0
assi (resi 148 and name HN ) (resi 148 and name HA ) 2.4 2.4 0.0
assi (resi 149 and name HN ) (resi 150 and name HN ) 2.9 2.9 0.0
assi (resi 149 and name HN ) (resi 149 and name HA ) 2.9 2.9 0.0
assi (resi 149 and name HN ) (resi 148 and name HA ) 3.4 3.4 0.0
assi (resi 150 and name HN ) (resi 149 and name HA ) 3.4 3.4 0.0
assi (resi 150 and name HN ) (resi 150 and name HA ) 2.9 2.9 0.0
assi (resi 150 and name HN ) (resi 151 and name HA ) 5.5 5.5 0.0
assi (resi 151 and name HN ) (resi 150 and name HN ) 2.4 2.4 0.0
assi (resi 151 and name HN ) (resi 150 and name HA ) 2.4 2.4 0.0
assi (resi 151 and name HN ) (resi 151 and name HA ) 2.9 2.9 0.0
assi (resi 152 and name HN ) (resi 153 and name HN ) 2.4 2.4 0.0
assi (resi 152 and name HN ) (resi 152 and name HA ) 2.9 2.9 0.0
assi (resi 153 and name HN ) (resi 153 and name HA ) 2.9 2.9 0.0
assi (resi 153 and name HN ) (resi 152 and name HA ) 3.4 3.4 0.0
assi (resi 154 and name HN ) (resi 154 and name HA ) 3.4 3.4 0.0
assi (resi 154 and name HN ) (resi 153 and name HA ) 2.4 2.4 0.0
assi (resi 155 and name HN ) (resi 154 and name HA ) 2.4 2.4 0.0
assi (resi 156 and name HN ) (resi 157 and name HN ) 4.0 4.0 0.0
assi (resi 156 and name HN ) (resi 156 and name HA ) 2.9 2.9 0.0
assi (resi 156 and name HN ) (resi 155 and name HA ) 2.4 2.4 0.0
assi (resi 157 and name HN ) (resi 158 and name HN ) 5.5 5.5 0.0
assi (resi 157 and name HN ) (resi 156 and name HA ) 2.9 2.9 0.0
assi (resi 157 and name HN ) (resi 157 and name HA ) 2.9 2.9 0.0
assi (resi 158 and name HN ) (resi 159 and name HN ) 5.5 5.5 0.0
assi (resi 158 and name HN ) (resi 157 and name HN ) 5.5 5.5 0.0
assi (resi 158 and name HN ) (resi 157 and name HA ) 2.4 2.4 0.0
assi (resi 159 and name HN ) (resi 160 and name HN ) 3.4 3.4 0.0
assi (resi 159 and name HN ) (resi 158 and name HA ) 2.9 2.9 0.0
assi (resi 159 and name HN ) (resi 159 and name HA ) 2.9 2.9 0.0
assi (resi 160 and name HN ) (resi 161 and name HN ) 3.4 3.4 0.0
assi (resi 160 and name HN ) (resi 160 and name HA ) 2.9 2.9 0.0
assi (resi 161 and name HN ) (resi 162 and name HN ) 2.9 2.9 0.0
assi (resi 161 and name HN ) (resi 161 and name HA ) 2.9 2.9 0.0
assi (resi 161 and name HN ) (resi 160 and name HA ) 5.5 5.5 0.0
assi (resi 162 and name HN ) (resi 161 and name HA ) 3.4 3.4 0.0
assi (resi 162 and name HN ) (resi 162 and name HA ) 2.4 2.4 0.0
assi (resi 163 and name HN ) (resi 162 and name HN ) 2.9 2.9 0.0
assi (resi 163 and name HN ) (resi 163 and name HA ) 2.9 2.9 0.0
assi (resi 163 and name HN ) (resi 162 and name HA ) 2.9 2.9 0.0
assi (resi 164 and name HN ) (resi 163 and name HN ) 4.0 4.0 0.0
assi (resi 164 and name HN ) (resi 165 and name HN ) 2.9 2.9 0.0
assi (resi 164 and name HN ) (resi 164 and name HA# ) 2.4 2.4 1.0
assi (resi 164 and name HN ) (resi 164 and name HA# ) 2.4 2.4 1.0
assi (resi 165 and name HN ) (resi 166 and name HN ) 2.9 2.9 0.0
assi (resi 165 and name HN ) (resi 164 and name HA# ) 2.9 2.9 1.0
assi (resi 165 and name HN ) (resi 165 and name HA ) 2.9 2.9 0.0
assi (resi 166 and name HN ) (resi 167 and name HN ) 2.9 2.9 0.0
assi (resi 166 and name HN ) (resi 166 and name HA ) 2.9 2.9 0.0
assi (resi 166 and name HN ) (resi 165 and name HA ) 3.4 3.4 0.0
assi (resi 167 and name HN ) (resi 167 and name HA ) 2.9 2.9 0.0
assi (resi 168 and name HN ) (resi 167 and name HN ) 2.9 2.9 0.0
assi (resi 168 and name HN ) (resi 167 and name HA ) 3.4 3.4 0.0
assi (resi 168 and name HN ) (resi 168 and name HA ) 2.9 2.9 0.0
assi (resi 169 and name HN ) (resi 168 and name HN ) 2.9 2.9 0.0
assi (resi 169 and name HN ) (resi 169 and name HA ) 2.9 2.9 0.0
assi (resi 170 and name HN ) (resi 169 and name HN ) 2.9 2.9 0.0
assi (resi 171 and name HN ) (resi 171 and name HA ) 2.9 2.9 0.0
assi (resi 172 and name HN ) (resi 173 and name HN ) 2.9 2.9 0.0
assi (resi 172 and name HN ) (resi 172 and name HA ) 2.9 2.9 0.0
assi (resi 173 and name HA ) (resi 173 and name HN ) 3.4 3.4 0.0
assi (resi 174 and name HN ) (resi 174 and name HA ) 2.9 2.9 0.0
assi (resi 174 and name HN ) (resi 173 and name HA ) 3.4 3.4 0.0
assi (resi 175 and name HN ) (resi 174 and name HN ) 2.4 2.4 0.0
assi (resi 175 and name HN ) (resi 176 and name HN ) 2.4 2.4 0.0
assi (resi 175 and name HN ) (resi 174 and name HA ) 2.9 2.9 0.0
assi (resi 175 and name HA ) (resi 175 and name HN ) 3.4 3.4 0.0
assi (resi 176 and name HN ) (resi 176 and name HA ) 2.9 2.9 0.0
assi (resi 177 and name HN ) (resi 177 and name HA ) 2.9 2.9 0.0
assi (resi 177 and name HN ) (resi 176 and name HA ) 2.9 2.9 0.0
assi (resi 178 and name HN ) (resi 178 and name HA ) 2.9 2.9 0.0
assi (resi 178 and name HN ) (resi 177 and name HA ) 2.9 2.9 0.0
assi (resi 179 and name HB# ) (resi 179 and name HA ) 3.4 3.4 1.0
assi (resi 181 and name HN ) (resi 180 and name HA ) 2.4 2.4 0.0
assi (resi 183 and name HN ) (resi 182 and name HA ) 2.4 2.4 0.0
assi (resi 184 and name HN ) (resi 183 and name HA ) 2.4 2.4 0.0
assi (resi 185 and name HN ) (resi 184 and name HA ) 2.4 2.4 0.0
assi (resi 186 and name HN ) (resi 185 and name HA ) 2.9 2.9 0.0
assi (resi 187 and name HN ) (resi 186 and name HA ) 2.4 2.4 0.0
assi (resi 187 and name HN ) (resi 187 and name HA ) 4.0 4.0 0.0

!INTRARESIDUE AND SEQUENTIAL HN OR HA TO SIDECHAIN

assi (resi  -6 and name HN ) (resi  -6 and name HB# ) 4.0 4.0 0.6
assi (resi  -5 and name HN ) (resi  -6 and name HB# ) 2.9 2.9 1.0
assi (resi  -5 and name HN ) (resi  -5 and name HB# ) 2.4 2.4 0.6
assi (resi  -5 and name HN ) (resi  -5 and name HG# ) 2.9 2.9 1.0
assi (resi  -4 and name HN ) (resi  -4 and name HB ) 2.9 2.9 0.0
assi (resi  -4 and name HN ) (resi  -4 and name HG1# ) 2.9 2.9 1.0
assi (resi  -4 and name HN ) (resi  -4 and name HG1# ) 2.9 2.9 1.0
assi (resi  -4 and name HA ) (resi  -4 and name HB ) 3.4 3.4 0.0
assi (resi  -4 and name HA ) (resi  -4 and name HG1# ) 4.0 4.0 0.6
assi (resi  -4 and name HA ) (resi  -4 and name HG1# ) 2.9 2.9 0.6
assi (resi  -4 and name HA ) (resi  -4 and name HG2# ) 2.9 2.9 1.1
assi (resi  -4 and name HA ) (resi  -4 and name HD# ) 2.9 2.9 0.0
assi (resi  -3 and name HN ) (resi  -3 and name HB ) 2.4 2.4 0.0
assi (resi  -3 and name HN ) (resi  -3 and name HG1# ) 2.4 2.4 1.0
assi (resi  -3 and name HN ) (resi  -3 and name HG1# ) 2.4 2.4 1.0
assi (resi  -2 and name HN ) (resi  -2 and name HB# ) 2.4 2.4 0.6
assi (resi  -2 and name HN ) (resi  -3 and name HB ) 3.4 3.4 0.0
assi (resi  -2 and name HN ) (resi  -3 and name HG1# ) 4.0 4.0 1.0
assi (resi  -2 and name HN ) (resi  -3 and name HG1# ) 4.0 4.0 1.0
assi (resi  -2 and name HN ) (resi  -3 and name HG2# ) 2.4 2.4 1.5
assi (resi  -1 and name HN ) (resi  -2 and name HB# ) 2.4 2.4 1.0
assi (resi  -1 and name HN ) (resi  -1 and name HB# ) 2.4 2.4 1.1
assi (resi  -1 and name HA ) (resi  -1 and name HB# ) 2.9 2.9 1.5
assi (resi   1 and name HN ) (resi   1 and name HG# ) 2.4 2.4 1.0
assi (resi   1 and name HN ) (resi   1 and name HB# ) 2.4 2.4 0.6
assi (resi   1 and name HN ) (resi  -1 and name HB# ) 2.4 2.4 1.5
assi (resi   1 and name HB# ) (resi   1 and name HA ) 3.4 3.4 1.0
assi (resi   1 and name HB# ) (resi   1 and name HA ) 2.9 2.9 1.0
assi (resi   1 and name HG# ) (resi   1 and name HA ) 2.9 2.9 0.6
assi (resi   2 and name HN ) (resi   2 and name HG# ) 2.4 2.4 1.0
assi (resi   2 and name HN ) (resi   2 and name HB# ) 2.4 2.4 0.6
assi (resi   2 and name HE2# ) (resi   2 and name HA ) 4.0 4.0 0.0
assi (resi   2 and name HE2# ) (resi   2 and name HA ) 3.4 3.4 0.0
assi (resi   4 and name HN ) (resi   4 and name HB ) 2.9 2.9 0.0
assi (resi   4 and name HN ) (resi   4 and name HG1# ) 3.4 3.4 1.0
assi (resi   4 and name HN ) (resi   4 and name HD# ) 2.9 2.9 0.0
assi (resi   4 and name HN ) (resi   4 and name HG2# ) 3.4 3.4 1.5
assi (resi   4 and name HA ) (resi   4 and name HB ) 5.5 5.5 0.0
assi (resi   4 and name HG2# ) (resi   4 and name HA ) 4.0 4.0 1.1
assi (resi   4 and name HD# ) (resi   5 and name HN ) 5.5 5.5 0.0
assi (resi   5 and name HN ) (resi   4 and name HG2# ) 2.9 2.9 1.5
assi (resi   6 and name HN ) (resi   6 and name HB# ) 3.4 3.4 0.6
assi (resi   6 and name HN ) (resi   6 and name HB# ) 3.4 3.4 0.6
assi (resi   7 and name HN ) (resi   6 and name HB# ) 3.4 3.4 1.0
assi (resi   7 and name HN ) (resi   6 and name HB# ) 5.5 5.5 1.0
assi (resi   7 and name HN ) (resi   7 and name HB ) 3.4 3.4 0.0
assi (resi   7 and name HN ) (resi   7 and name HG2# ) 2.9 2.9 1.5
assi (resi   7 and name HN ) (resi   7 and name HG1# ) 2.4 2.4 1.5
assi (resi   7 and name HA ) (resi   7 and name HG2# ) 3.4 3.4 1.1
assi (resi   7 and name HA ) (resi   7 and name HG1# ) 4.0 4.0 1.1
assi (resi   8 and name HN ) (resi   8 and name HB ) 2.9 2.9 0.0
assi (resi   8 and name HN ) (resi   7 and name HB ) 5.5 5.5 0.0
assi (resi   8 and name HN ) (resi   7 and name HG1# ) 2.9 2.9 1.5
assi (resi   8 and name HG2# ) (resi   8 and name HN ) 3.4 3.4 1.5
assi (resi   8 and name HG2# ) (resi   8 and name HA ) 3.4 3.4 1.1
assi (resi   8 and name HG1# ) (resi   9 and name HN ) 2.9 2.9 1.5
assi (resi   8 and name HG1# ) (resi   8 and name HA ) 5.5 5.5 1.1
assi (resi   9 and name HN ) (resi   9 and name HG1# ) 2.4 2.4 1.5
assi (resi   9 and name HA ) (resi   9 and name HB ) 3.4 3.4 0.0
assi (resi   9 and name HA ) (resi   8 and name HG1# ) 3.4 3.4 1.5
assi (resi   9 and name HG1# ) (resi   9 and name HN ) 5.5 5.5 1.5
assi (resi   9 and name HG2# ) (resi   9 and name HN ) 3.4 3.4 1.5
assi (resi  10 and name HN ) (resi   9 and name HG1# ) 3.4 3.4 1.5
assi (resi  10 and name HN ) (resi   9 and name HG2# ) 3.4 3.4 1.5
assi (resi  11 and name HN ) (resi  11 and name HB# ) 3.4 3.4 0.6
assi (resi  12 and name HN ) (resi  11 and name HB# ) 2.9 2.9 1.0
assi (resi  12 and name HN ) (resi  13 and name HB# ) 5.5 5.5 1.5
assi (resi  13 and name HN ) (resi  13 and name HB# ) 2.9 2.9 1.1
assi (resi  13 and name HA ) (resi  13 and name HB# ) 2.9 2.9 1.5
assi (resi  13 and name HB# ) (resi  12 and name HA# ) 3.4 3.4 2.5
assi (resi  14 and name HN ) (resi  14 and name HB ) 3.4 3.4 0.0
assi (resi  17 and name HN ) (resi  17 and name HG2# ) 2.9 2.9 1.5
assi (resi  17 and name HA ) (resi  17 and name HG2# ) 4.0 4.0 1.1
assi (resi  18 and name HN ) (resi  17 and name HG2# ) 4.0 4.0 1.5
assi (resi  19 and name HN ) (resi  19 and name HB# ) 3.4 3.4 0.6
assi (resi  19 and name HB# ) (resi  19 and name HA ) 3.4 3.4 1.0
assi (resi  19 and name HB# ) (resi  19 and name HN ) 4.0 4.0 0.6
assi (resi  19 and name HB# ) (resi  19 and name HA ) 3.4 3.4 1.0
assi (resi  19 and name HD1# ) (resi  19 and name HN ) 4.0 4.0 1.0
assi (resi  19 and name HD1# ) (resi  20 and name HN ) 4.0 4.0 1.0
assi (resi  19 and name HD1# ) (resi  19 and name HA ) 2.9 2.9 1.0
assi (resi  19 and name HD2# ) (resi  19 and name HN ) 4.0 4.0 1.0
assi (resi  19 and name HD2# ) (resi  20 and name HN ) 4.0 4.0 1.0
assi (resi  19 and name HD2# ) (resi  19 and name HA ) 3.4 3.4 1.0
assi (resi  20 and name HN ) (resi  19 and name HB# ) 3.4 3.4 1.0
assi (resi  20 and name HN ) (resi  19 and name HG ) 2.9 2.9 0.0
assi (resi  20 and name HN ) (resi  19 and name HB# ) 2.9 2.9 1.0
assi (resi  20 and name HD1# ) (resi  20 and name HA ) 5.5 5.5 1.0
assi (resi  21 and name HN ) (resi  21 and name HB ) 2.9 2.9 0.0
assi (resi  21 and name HA ) (resi  21 and name HB ) 4.0 4.0 0.0
assi (resi  21 and name HG2# ) (resi  21 and name HN ) 4.0 4.0 1.5
assi (resi  21 and name HG2# ) (resi  21 and name HA ) 3.4 3.4 1.1
assi (resi  21 and name HD# ) (resi  21 and name HN ) 4.0 4.0 0.0
assi (resi  21 and name HD# ) (resi  21 and name HA ) 2.9 2.9 0.0
assi (resi  22 and name HN ) (resi  21 and name HG1# ) 2.9 2.9 1.0
assi (resi  22 and name HB# ) (resi  22 and name HN ) 3.4 3.4 0.6
assi (resi  23 and name HN ) (resi  23 and name HD# ) 2.9 2.9 2.0
assi (resi  23 and name HN ) (resi  22 and name HB# ) 2.9 2.9 1.0
assi (resi  23 and name HN ) (resi  23 and name HB# ) 2.9 2.9 0.6
assi (resi  23 and name HN ) (resi  23 and name HB# ) 3.4 3.4 0.6
assi (resi  23 and name HA ) (resi  23 and name HD# ) 4.0 4.0 1.0
assi (resi  23 and name HA ) (resi  23 and name HB# ) 4.0 4.0 1.0
assi (resi  24 and name HN ) (resi  23 and name HD# ) 3.4 3.4 2.0
assi (resi  24 and name HN ) (resi  24 and name HB ) 4.0 4.0 0.0
assi (resi  24 and name HN ) (resi  23 and name HB# ) 4.0 4.0 1.0
assi (resi  24 and name HN ) (resi  23 and name HB# ) 5.5 5.5 1.0
assi (resi  24 and name HN ) (resi  24 and name HG2# ) 2.4 2.4 1.5
assi (resi  24 and name HA ) (resi  24 and name HG2# ) 2.4 2.4 1.1
assi (resi  25 and name HN ) (resi  25 and name HG2# ) 2.4 2.4 1.5
assi (resi  25 and name HA ) (resi  25 and name HB ) 3.4 3.4 0.0
assi (resi  25 and name HA ) (resi  25 and name HG2# ) 2.4 2.4 1.1
assi (resi  26 and name HN ) (resi  26 and name HB# ) 2.9 2.9 0.6
assi (resi  26 and name HN ) (resi  26 and name HB# ) 2.9 2.9 0.6
assi (resi  26 and name HN ) (resi  27 and name HB# ) 5.5 5.5 1.0
assi (resi  26 and name HN ) (resi  25 and name HG2# ) 3.4 3.4 1.5
assi (resi  26 and name HD2# ) (resi  26 and name HA ) 3.4 3.4 0.0
assi (resi  26 and name HD2# ) (resi  26 and name HA ) 3.4 3.4 0.0
assi (resi  26 and name HB# ) (resi  26 and name HA ) 4.0 4.0 1.0
assi (resi  26 and name HB# ) (resi  26 and name HA ) 4.0 4.0 1.0
assi (resi  27 and name HN ) (resi  26 and name HB# ) 2.9 2.9 1.0
assi (resi  27 and name HN ) (resi  27 and name HB# ) 2.4 2.4 0.6
assi (resi  27 and name HA ) (resi  27 and name HB# ) 2.4 2.4 1.0
assi (resi  27 and name HA ) (resi  27 and name HB# ) 2.4 2.4 1.0
assi (resi  27 and name HA ) (resi  27 and name HG# ) 2.4 2.4 0.6
assi (resi  27 and name HB# ) (resi  27 and name HA ) 3.4 3.4 1.0
assi (resi  27 and name HB# ) (resi  27 and name HA ) 3.4 3.4 1.0
assi (resi  27 and name HG# ) (resi  27 and name HA ) 2.9 2.9 0.6
assi (resi  27 and name HD# ) (resi  27 and name HA ) 2.9 2.9 1.0
assi (resi  28 and name HN ) (resi  28 and name HD# ) 2.9 2.9 2.0
assi (resi  28 and name HN ) (resi  28 and name HB# ) 2.4 2.4 0.6
assi (resi  28 and name HN ) (resi  28 and name HB# ) 2.4 2.4 0.6
assi (resi  28 and name HN ) (resi  27 and name HG# ) 2.9 2.9 1.0
assi (resi  29 and name HA ) (resi  28 and name HD# ) 4.0 4.0 2.0
assi (resi  30 and name HN ) (resi  30 and name HB# ) 2.4 2.4 0.6
assi (resi  30 and name HA ) (resi  30 and name HB# ) 2.9 2.9 1.0
assi (resi  30 and name HA ) (resi  30 and name HB# ) 2.4 2.4 1.0
assi (resi  31 and name HN ) (resi  30 and name HB# ) 3.4 3.4 1.0
assi (resi  31 and name HN ) (resi  30 and name HB# ) 2.9 2.9 1.0
assi (resi  31 and name HN ) (resi  32 and name HB# ) 5.5 5.5 1.0
assi (resi  31 and name HN ) (resi  31 and name HG# ) 2.4 2.4 1.0
assi (resi  31 and name HN ) (resi  31 and name HB# ) 2.4 2.4 0.6
assi (resi  31 and name HN ) (resi  31 and name HB# ) 2.4 2.4 0.6
assi (resi  31 and name HB# ) (resi  31 and name HA ) 2.9 2.9 1.0
assi (resi  31 and name HB# ) (resi  31 and name HA ) 2.4 2.4 1.0
assi (resi  31 and name HG# ) (resi  31 and name HA ) 2.4 2.4 0.6
assi (resi  32 and name HN ) (resi  32 and name HD# ) 3.4 3.4 2.0
assi (resi  32 and name HN ) (resi  32 and name HB# ) 2.4 2.4 0.6
assi (resi  32 and name HN ) (resi  32 and name HB# ) 2.4 2.4 0.6
assi (resi  32 and name HN ) (resi  33 and name HG1# ) 4.0 4.0 1.5
assi (resi  32 and name HA ) (resi  32 and name HD# ) 2.9 2.9 1.0
assi (resi  32 and name HA ) (resi  32 and name HB# ) 2.9 2.9 1.0
assi (resi  32 and name HA ) (resi  32 and name HB# ) 2.9 2.9 1.0
assi (resi  32 and name HB# ) (resi  32 and name HA ) 3.4 3.4 1.0
assi (resi  33 and name HN ) (resi  32 and name HD# ) 2.9 2.9 2.0
assi (resi  33 and name HN ) (resi  32 and name HE# ) 3.4 3.4 2.0
assi (resi  33 and name HN ) (resi  34 and name HD# ) 4.0 4.0 1.0
assi (resi  33 and name HN ) (resi  32 and name HB# ) 3.4 3.4 1.0
assi (resi  33 and name HN ) (resi  32 and name HB# ) 4.0 4.0 1.0
assi (resi  33 and name HN ) (resi  33 and name HG1# ) 2.4 2.4 1.5
assi (resi  33 and name HA ) (resi  34 and name HD# ) 2.9 2.9 1.0
assi (resi  33 and name HA ) (resi  33 and name HG2# ) 3.4 3.4 1.1
assi (resi  33 and name HB ) (resi  33 and name HA ) 4.0 4.0 0.0
assi (resi  33 and name HG1# ) (resi  33 and name HA ) 2.9 2.9 1.1
assi (resi  35 and name HN ) (resi  35 and name HG2# ) 4.0 4.0 1.5
assi (resi  35 and name HN ) (resi  36 and name HG## ) 4.0 4.0 2.9
assi (resi  35 and name HA ) (resi  35 and name HG2# ) 4.0 4.0 1.1
assi (resi  36 and name HN ) (resi  35 and name HG2# ) 3.4 3.4 1.5
assi (resi  36 and name HN ) (resi  36 and name HG## ) 4.0 4.0 2.2
assi (resi  36 and name HN ) (resi  36 and name HG## ) 4.0 4.0 2.2
assi (resi  39 and name HD2# ) (resi  39 and name HA ) 4.0 4.0 0.0
assi (resi  39 and name HD2# ) (resi  39 and name HA ) 5.5 5.5 0.0
assi (resi  41 and name HB# ) (resi  41 and name HA ) 2.9 2.9 1.5
assi (resi  42 and name HN ) (resi  42 and name HB ) 3.4 3.4 0.0
assi (resi  42 and name HN ) (resi  42 and name HG## ) 2.4 2.4 2.2
assi (resi  42 and name HA ) (resi  42 and name HB ) 2.9 2.9 0.0
assi (resi  42 and name HG2# ) (resi  42 and name HA ) 4.0 4.0 1.1
assi (resi  43 and name HN ) (resi  43 and name HB ) 2.4 2.4 0.0
assi (resi  43 and name HN ) (resi  42 and name HB ) 2.4 2.4 0.0
assi (resi  43 and name HN ) (resi  42 and name HG2# ) 2.4 2.4 1.5
assi (resi  43 and name HA ) (resi  43 and name HB ) 4.0 4.0 0.0
assi (resi  43 and name HA ) (resi  43 and name HG2# ) 2.4 2.4 1.1
assi (resi  43 and name HG2# ) (resi  44 and name HN ) 4.0 4.0 1.5
assi (resi  44 and name HN ) (resi  43 and name HB ) 4.0 4.0 0.0
assi (resi  44 and name HN ) (resi  44 and name HB ) 3.4 3.4 0.0
assi (resi  44 and name HN ) (resi  44 and name HG## ) 2.4 2.4 2.2
assi (resi  44 and name HA ) (resi  44 and name HB ) 3.4 3.4 0.0
assi (resi  44 and name HG1# ) (resi  44 and name HA ) 2.9 2.9 1.1
assi (resi  44 and name HG2# ) (resi  44 and name HA ) 2.9 2.9 1.1
assi (resi  45 and name HN ) (resi  45 and name HG# ) 2.4 2.4 1.0
assi (resi  45 and name HN ) (resi  44 and name HB ) 2.4 2.4 0.0
assi (resi  45 and name HN ) (resi  44 and name HG## ) 2.4 2.4 2.9
assi (resi  45 and name HB# ) (resi  45 and name HA ) 4.0 4.0 1.0
assi (resi  45 and name HG# ) (resi  45 and name HA ) 5.5 5.5 0.6
assi (resi  45 and name HG# ) (resi  45 and name HA ) 4.0 4.0 0.6
assi (resi  45 and name HE# ) (resi  45 and name HA ) 2.9 2.9 1.5
assi (resi  46 and name HN ) (resi  45 and name HG# ) 5.5 5.5 1.0
assi (resi  46 and name HN ) (resi  46 and name HB ) 2.4 2.4 0.0
assi (resi  46 and name HN ) (resi  46 and name HG1# ) 2.9 2.9 1.0
assi (resi  46 and name HN ) (resi  46 and name HG2# ) 2.9 2.9 1.5
assi (resi  46 and name HN ) (resi  46 and name HD# ) 2.9 2.9 0.0
assi (resi  46 and name HA ) (resi  46 and name HB ) 3.4 3.4 0.0
assi (resi  46 and name HA ) (resi  46 and name HG1# ) 2.9 2.9 0.6
assi (resi  46 and name HG2# ) (resi  46 and name HA ) 2.9 2.9 1.1
assi (resi  46 and name HG2# ) (resi  47 and name HA# ) 3.4 3.4 2.5
assi (resi  46 and name HD# ) (resi  46 and name HA ) 2.9 2.9 0.0
assi (resi  47 and name HN ) (resi  46 and name HB ) 3.4 3.4 0.0
assi (resi  47 and name HN ) (resi  46 and name HG1# ) 4.0 4.0 1.0
assi (resi  47 and name HN ) (resi  46 and name HG2# ) 2.4 2.4 1.5
assi (resi  49 and name HN ) (resi  49 and name HG# ) 2.9 2.9 1.0
assi (resi  49 and name HN ) (resi  49 and name HB# ) 2.4 2.4 0.6
assi (resi  49 and name HA ) (resi  49 and name HB# ) 2.9 2.9 1.0
assi (resi  49 and name HG# ) (resi  49 and name HA ) 5.5 5.5 0.6
assi (resi  49 and name HG# ) (resi  49 and name HA ) 4.0 4.0 0.6
assi (resi  51 and name HN ) (resi  51 and name HD# ) 2.9 2.9 2.0
assi (resi  51 and name HN ) (resi  51 and name HB# ) 2.9 2.9 0.6
assi (resi  51 and name HN ) (resi  50 and name HB# ) 2.4 2.4 1.0
assi (resi  52 and name HN ) (resi  51 and name HD# ) 4.0 4.0 2.0
assi (resi  52 and name HN ) (resi  52 and name HB ) 2.9 2.9 0.0
assi (resi  52 and name HN ) (resi  51 and name HB# ) 2.9 2.9 1.0
assi (resi  52 and name HN ) (resi  52 and name HG2# ) 2.4 2.4 1.5
assi (resi  52 and name HB ) (resi  53 and name HA ) 4.0 4.0 0.0
assi (resi  52 and name HG2# ) (resi  53 and name HN ) 3.4 3.4 1.5
assi (resi  52 and name HG2# ) (resi  53 and name HA ) 2.9 2.9 1.5
assi (resi  53 and name HN ) (resi  53 and name HB# ) 2.9 2.9 0.6
assi (resi  53 and name HN ) (resi  53 and name HD2# ) 2.9 2.9 1.0
assi (resi  53 and name HA ) (resi  53 and name HD1# ) 2.9 2.9 1.0
assi (resi  53 and name HD2# ) (resi  53 and name HA ) 5.5 5.5 1.0
assi (resi  55 and name HN ) (resi  55 and name HD## ) 2.9 2.9 2.9
assi (resi  55 and name HN ) (resi  55 and name HD## ) 2.9 2.9 2.9
assi (resi  55 and name HD2# ) (resi  55 and name HA ) 5.5 5.5 1.0
assi (resi  56 and name HN ) (resi  56 and name HD# ) 2.9 2.9 2.0
assi (resi  56 and name HN ) (resi  56 and name HB# ) 3.4 3.4 0.6
assi (resi  56 and name HN ) (resi  55 and name HD## ) 2.9 2.9 2.9
assi (resi  57 and name HN ) (resi  56 and name HD# ) 3.4 3.4 2.0
assi (resi  62 and name HN ) (resi  61 and name HG# ) 3.4 3.4 1.0
assi (resi  62 and name HN ) (resi  62 and name HG# ) 2.4 2.4 1.0
assi (resi  62 and name HN ) (resi  62 and name HB# ) 2.4 2.4 0.6
assi (resi  62 and name HA ) (resi  62 and name HG# ) 2.9 2.9 0.6
assi (resi  62 and name HA ) (resi  62 and name HB# ) 2.4 2.4 1.0
assi (resi  63 and name HN ) (resi  64 and name HD# ) 5.5 5.5 2.0
assi (resi  63 and name HN ) (resi  63 and name HB# ) 2.4 2.4 0.6
assi (resi  63 and name HN ) (resi  62 and name HB# ) 2.4 2.4 1.0
assi (resi  63 and name HB# ) (resi  63 and name HA ) 3.4 3.4 1.0
assi (resi  63 and name HB# ) (resi  63 and name HA ) 3.4 3.4 1.0
assi (resi  64 and name HN ) (resi  64 and name HD# ) 2.9 2.9 2.0
assi (resi  64 and name HN ) (resi  64 and name HB# ) 3.4 3.4 0.6
assi (resi  64 and name HN ) (resi  64 and name HB# ) 2.9 2.9 0.6
assi (resi  64 and name HN ) (resi  63 and name HB# ) 2.9 2.9 1.0
assi (resi  64 and name HA ) (resi  64 and name HD# ) 4.0 4.0 1.0
assi (resi  64 and name HA ) (resi  64 and name HB# ) 4.0 4.0 1.0
assi (resi  64 and name HA ) (resi  64 and name HB# ) 4.0 4.0 1.0
assi (resi  65 and name HN ) (resi  64 and name HD# ) 4.0 4.0 2.0
assi (resi  70 and name HN ) (resi  70 and name HG ) 2.9 2.9 0.0
assi (resi  70 and name HN ) (resi  70 and name HD2# ) 2.4 2.4 1.0
assi (resi  70 and name HA ) (resi  70 and name HB# ) 2.9 2.9 1.0
assi (resi  70 and name HA ) (resi  70 and name HB# ) 3.4 3.4 1.0
assi (resi  70 and name HG ) (resi  70 and name HA ) 5.5 5.5 0.0
assi (resi  70 and name HD1# ) (resi  70 and name HA ) 2.9 2.9 1.0
assi (resi  70 and name HD2# ) (resi  70 and name HA ) 2.4 2.4 1.0
assi (resi  71 and name HA ) (resi  71 and name HB# ) 3.4 3.4 1.0
assi (resi  72 and name HN ) (resi  72 and name HB# ) 2.9 2.9 0.6
assi (resi  75 and name HN ) (resi  75 and name HG2# ) 2.9 2.9 1.5
assi (resi  75 and name HG2# ) (resi  74 and name HA ) 3.4 3.4 1.5
assi (resi  76 and name HN ) (resi  77 and name HG2# ) 3.4 3.4 1.5
assi (resi  76 and name HA ) (resi  77 and name HB ) 5.5 5.5 0.0
assi (resi  77 and name HN ) (resi  77 and name HB ) 3.4 3.4 0.0
assi (resi  77 and name HN ) (resi  77 and name HG1# ) 2.4 2.4 1.5
assi (resi  77 and name HN ) (resi  77 and name HG2# ) 2.4 2.4 1.5
assi (resi  77 and name HG1# ) (resi  78 and name HA ) 4.0 4.0 1.5
assi (resi  77 and name HG2# ) (resi  78 and name HA ) 5.5 5.5 1.5
assi (resi  78 and name HN ) (resi  78 and name HD# ) 2.9 2.9 2.0
assi (resi  78 and name HN ) (resi  77 and name HG1# ) 3.4 3.4 1.5
assi (resi  79 and name HN ) (resi  78 and name HD# ) 4.0 4.0 2.0
assi (resi  79 and name HN ) (resi  78 and name HB# ) 3.4 3.4 1.0
assi (resi  79 and name HN ) (resi  79 and name HG ) 4.0 4.0 0.0
assi (resi  79 and name HD1# ) (resi  79 and name HN ) 5.5 5.5 1.0
assi (resi  79 and name HD2# ) (resi  79 and name HN ) 3.4 3.4 1.0
assi (resi  80 and name HN ) (resi  80 and name HB ) 2.9 2.9 0.0
assi (resi  81 and name HN ) (resi  81 and name HB# ) 2.9 2.9 0.6
assi (resi  81 and name HN ) (resi  81 and name HB# ) 3.4 3.4 0.6
assi (resi  81 and name HN ) (resi  80 and name HG1# ) 2.4 2.4 1.5
assi (resi  82 and name HN ) (resi  82 and name HD# ) 2.9 2.9 2.0
assi (resi  82 and name HN ) (resi  81 and name HB# ) 3.4 3.4 1.0
assi (resi  82 and name HA ) (resi  82 and name HB# ) 2.9 2.9 1.0
assi (resi  82 and name HA ) (resi  82 and name HB# ) 2.9 2.9 1.0
assi (resi  83 and name HN ) (resi  82 and name HD# ) 3.4 3.4 2.0
assi (resi  83 and name HN ) (resi  83 and name HB# ) 3.4 3.4 0.6
assi (resi  83 and name HN ) (resi  84 and name HG2# ) 3.4 3.4 1.5
assi (resi  83 and name HA ) (resi  84 and name HG2# ) 4.0 4.0 1.5
assi (resi  84 and name HB ) (resi  84 and name HA ) 5.5 5.5 0.0
assi (resi  84 and name HG2# ) (resi  84 and name HN ) 3.4 3.4 1.5
assi (resi  84 and name HG2# ) (resi  85 and name HN ) 3.4 3.4 1.5
assi (resi  84 and name HG2# ) (resi  84 and name HB ) 3.4 3.4 1.5
assi (resi  84 and name HG1# ) (resi  84 and name HN ) 3.4 3.4 1.5
assi (resi  84 and name HG1# ) (resi  84 and name HA ) 3.4 3.4 1.1
assi (resi  85 and name HN ) (resi  85 and name HB ) 3.4 3.4 0.0
assi (resi  85 and name HN ) (resi  85 and name HG2# ) 2.4 2.4 1.5
assi (resi  85 and name HA ) (resi  85 and name HG## ) 5.5 5.5 1.1
assi (resi  85 and name HG1# ) (resi  86 and name HN ) 4.0 4.0 1.5
assi (resi  85 and name HG1# ) (resi  85 and name HA ) 2.9 2.9 1.1
assi (resi  85 and name HG2# ) (resi  84 and name HN ) 5.5 5.5 1.5
assi (resi  85 and name HG2# ) (resi  86 and name HN ) 2.9 2.9 1.5
assi (resi  85 and name HG2# ) (resi  85 and name HA ) 4.0 4.0 1.1
assi (resi  86 and name HN ) (resi  86 and name HB# ) 3.4 3.4 0.6
assi (resi  86 and name HN ) (resi  86 and name HB# ) 2.9 2.9 0.6
assi (resi  86 and name HA ) (resi  87 and name HD# ) 5.5 5.5 1.0
assi (resi  88 and name HN ) (resi  88 and name HB# ) 2.9 2.9 0.6
assi (resi  88 and name HN ) (resi  88 and name HB# ) 2.4 2.4 0.6
assi (resi  89 and name HN ) (resi  90 and name HD# ) 3.4 3.4 2.0
assi (resi  89 and name HN ) (resi  88 and name HB# ) 2.9 2.9 1.0
assi (resi  89 and name HA ) (resi  89 and name HB# ) 3.4 3.4 1.0
assi (resi  90 and name HN ) (resi  90 and name HD# ) 2.9 2.9 2.0
assi (resi  90 and name HN ) (resi  90 and name HB# ) 3.4 3.4 0.6
assi (resi  90 and name HN ) (resi  90 and name HB# ) 3.4 3.4 0.6
assi (resi  90 and name HN ) (resi  91 and name HB# ) 5.5 5.5 1.0
assi (resi  90 and name HA ) (resi  90 and name HD# ) 4.0 4.0 1.0
assi (resi  90 and name HA ) (resi  90 and name HB# ) 5.5 5.5 1.0
assi (resi  91 and name HN ) (resi  90 and name HD# ) 3.4 3.4 2.0
assi (resi  92 and name HN ) (resi  92 and name HB# ) 3.4 3.4 0.6
assi (resi  92 and name HN ) (resi  92 and name HB# ) 3.4 3.4 0.6
assi (resi  92 and name HN ) (resi  91 and name HB# ) 2.4 2.4 1.0
assi (resi  92 and name HD2# ) (resi  92 and name HN ) 3.4 3.4 0.0
assi (resi  92 and name HD2# ) (resi  92 and name HN ) 3.4 3.4 0.0
assi (resi  93 and name HN ) (resi  93 and name HB ) 2.4 2.4 0.0
assi (resi  93 and name HN ) (resi  93 and name HG## ) 2.4 2.4 2.2
assi (resi  94 and name HN ) (resi  93 and name HB ) 2.9 2.9 0.0
assi (resi  94 and name HN ) (resi  94 and name HB# ) 2.9 2.9 0.6
assi (resi  94 and name HN ) (resi  94 and name HB# ) 2.9 2.9 0.6
assi (resi  94 and name HN ) (resi  94 and name HG# ) 2.9 2.9 1.0
assi (resi  94 and name HN ) (resi  93 and name HG1# ) 2.9 2.9 1.5
assi (resi  94 and name HA ) (resi  94 and name HG# ) 3.4 3.4 0.6
assi (resi  94 and name HB# ) (resi  94 and name HA ) 3.4 3.4 1.0
assi (resi  94 and name HB# ) (resi  94 and name HA ) 3.4 3.4 1.0
assi (resi  94 and name HD# ) (resi  94 and name HA ) 2.9 2.9 1.0
assi (resi  94 and name HE# ) (resi  94 and name HA ) 2.9 2.9 1.0
assi (resi  95 and name HN ) (resi  95 and name HG# ) 2.4 2.4 1.0
assi (resi  95 and name HN ) (resi  95 and name HB# ) 2.4 2.4 0.6
assi (resi  95 and name HN ) (resi  94 and name HG# ) 3.4 3.4 1.0
assi (resi  95 and name HB# ) (resi  95 and name HA ) 3.4 3.4 1.0
assi (resi  95 and name HG# ) (resi  95 and name HA ) 2.9 2.9 0.6
assi (resi  95 and name HG# ) (resi  95 and name HA ) 3.4 3.4 0.6
assi (resi  96 and name HN ) (resi  97 and name HD1 ) 3.4 3.4 0.0
assi (resi  96 and name HN ) (resi  96 and name HB# ) 2.9 2.9 0.6
assi (resi  97 and name HN ) (resi  97 and name HE1 ) 5.5 5.5 0.0
assi (resi  97 and name HN ) (resi  97 and name HD1 ) 3.4 3.4 0.0
assi (resi  97 and name HN ) (resi  97 and name HB# ) 3.4 3.4 0.6
assi (resi  97 and name HN ) (resi  97 and name HB# ) 2.9 2.9 0.6
assi (resi  98 and name HN ) (resi  97 and name HD1 ) 5.5 5.5 0.0
assi (resi  98 and name HN ) (resi  99 and name HD# ) 2.9 2.9 1.0
assi (resi  98 and name HN ) (resi  97 and name HB# ) 2.9 2.9 1.0
assi (resi  98 and name HN ) (resi  98 and name HB ) 2.9 2.9 0.0
assi (resi  98 and name HG2# ) (resi  98 and name HN ) 3.4 3.4 1.5
assi (resi  98 and name HG2# ) (resi  98 and name HA ) 2.9 2.9 1.1
assi (resi 100 and name HN ) (resi  99 and name HD# ) 3.4 3.4 1.0
assi (resi 100 and name HN ) (resi 100 and name HB# ) 2.4 2.4 0.6
assi (resi 100 and name HN ) (resi 101 and name HG1# ) 4.0 4.0 1.0
assi (resi 100 and name HN ) (resi 101 and name HG2# ) 3.4 3.4 1.5
assi (resi 101 and name HN ) (resi 100 and name HB# ) 3.4 3.4 1.0
assi (resi 101 and name HN ) (resi 101 and name HG1# ) 2.9 2.9 1.0
assi (resi 101 and name HN ) (resi 101 and name HB ) 3.4 3.4 0.0
assi (resi 101 and name HN ) (resi 101 and name HG2# ) 2.9 2.9 1.5
assi (resi 101 and name HB ) (resi 101 and name HA ) 5.5 5.5 0.0
assi (resi 101 and name HG2# ) (resi 101 and name HN ) 3.4 3.4 1.5
assi (resi 101 and name HG2# ) (resi 101 and name HA ) 2.9 2.9 1.1
assi (resi 101 and name HD# ) (resi 101 and name HN ) 4.0 4.0 0.0
assi (resi 102 and name HN ) (resi 101 and name HG1# ) 2.9 2.9 1.0
assi (resi 102 and name HN ) (resi 101 and name HD# ) 3.4 3.4 0.0
assi (resi 102 and name HA ) (resi 102 and name HG2# ) 3.4 3.4 1.1
assi (resi 102 and name HG2# ) (resi 102 and name HN ) 3.4 3.4 1.5
assi (resi 102 and name HG2# ) (resi 101 and name HA ) 3.4 3.4 1.5
assi (resi 103 and name HN ) (resi 103 and name HB# ) 2.4 2.4 0.6
assi (resi 103 and name HA ) (resi 103 and name HB# ) 2.9 2.9 1.0
assi (resi 104 and name HN ) (resi 103 and name HB# ) 5.5 5.5 1.0
assi (resi 104 and name HN ) (resi 104 and name HB# ) 5.5 5.5 0.6
assi (resi 105 and name HN ) (resi 106 and name HD# ) 2.9 2.9 1.0
assi (resi 105 and name HN ) (resi 105 and name HB# ) 2.9 2.9 0.6
assi (resi 105 and name HA ) (resi 106 and name HD# ) 4.0 4.0 1.0
assi (resi 106 and name HA ) (resi 102 and name HG2# ) 4.0 4.0 1.5
assi (resi 107 and name HN ) (resi 106 and name HD# ) 3.4 3.4 1.0
assi (resi 107 and name HN ) (resi 107 and name HB# ) 2.4 2.4 0.6
assi (resi 107 and name HN ) (resi 107 and name HB# ) 2.4 2.4 0.6
assi (resi 107 and name HN ) (resi 107 and name HD# ) 2.9 2.9 1.0
assi (resi 107 and name HA ) (resi 107 and name HB# ) 3.4 3.4 1.0
assi (resi 107 and name HA ) (resi 107 and name HB# ) 3.4 3.4 1.0
assi (resi 107 and name HA ) (resi 107 and name HD# ) 3.4 3.4 1.0
assi (resi 108 and name HN ) (resi 108 and name HB ) 2.9 2.9 0.0
assi (resi 108 and name HN ) (resi 109 and name HD# ) 2.9 2.9 1.0
assi (resi 108 and name HN ) (resi 107 and name HB# ) 3.4 3.4 1.0
assi (resi 108 and name HN ) (resi 107 and name HB# ) 3.4 3.4 1.0
assi (resi 110 and name HN ) (resi 110 and name HD# ) 2.4 2.4 2.0
assi (resi 110 and name HN ) (resi 110 and name HB# ) 2.9 2.9 0.6
assi (resi 110 and name HN ) (resi 110 and name HB# ) 2.9 2.9 0.6
assi (resi 110 and name HN ) (resi 111 and name HB# ) 5.5 5.5 1.0
assi (resi 110 and name HN ) (resi 111 and name HD1# ) 4.0 4.0 1.0
assi (resi 110 and name HA ) (resi 111 and name HD1# ) 5.5 5.5 1.0
assi (resi 111 and name HN ) (resi 110 and name HD# ) 2.9 2.9 2.0
assi (resi 111 and name HN ) (resi 111 and name HB# ) 2.4 2.4 0.6
assi (resi 111 and name HN ) (resi 111 and name HD## ) 2.4 2.4 2.9
assi (resi 111 and name HD1# ) (resi 110 and name HN ) 3.4 3.4 1.0
assi (resi 111 and name HD2# ) (resi 112 and name HN ) 2.9 2.9 1.0
assi (resi 111 and name HD2# ) (resi 111 and name HA ) 5.5 5.5 1.0
assi (resi 112 and name HN ) (resi 112 and name HB# ) 2.9 2.9 0.6
assi (resi 112 and name HD1# ) (resi 113 and name HN ) 4.0 4.0 1.0
assi (resi 112 and name HD1# ) (resi 112 and name HA ) 3.4 3.4 1.0
assi (resi 112 and name HD2# ) (resi 112 and name HA ) 5.5 5.5 1.0
assi (resi 113 and name HA ) (resi 113 and name HG2# ) 3.4 3.4 1.1
assi (resi 113 and name HG2# ) (resi 113 and name HN ) 4.0 4.0 1.5
assi (resi 113 and name HG2# ) (resi 112 and name HA ) 5.5 5.5 1.5
assi (resi 113 and name HG1# ) (resi 113 and name HN ) 3.4 3.4 1.5
assi (resi 113 and name HG1# ) (resi 112 and name HA ) 5.5 5.5 1.5
assi (resi 114 and name HN ) (resi 113 and name HG## ) 2.9 2.9 2.9
assi (resi 115 and name HN ) (resi 115 and name HG2# ) 2.4 2.4 1.5
assi (resi 115 and name HG2# ) (resi 115 and name HA ) 5.5 5.5 1.1
assi (resi 117 and name HN ) (resi 117 and name HB ) 3.4 3.4 0.0
assi (resi 117 and name HA ) (resi 117 and name HB ) 5.5 5.5 0.0
assi (resi 117 and name HG2# ) (resi 117 and name HN ) 3.4 3.4 1.5
assi (resi 117 and name HD# ) (resi 117 and name HN ) 5.5 5.5 0.0
assi (resi 118 and name HN ) (resi 118 and name HB# ) 2.9 2.9 0.6
assi (resi 118 and name HN ) (resi 117 and name HG2# ) 2.9 2.9 1.5
assi (resi 119 and name HN ) (resi 118 and name HB# ) 3.4 3.4 1.0
assi (resi 119 and name HN ) (resi 119 and name HB# ) 2.4 2.4 0.6
assi (resi 119 and name HB# ) (resi 119 and name HA ) 4.0 4.0 1.0
assi (resi 119 and name HD1# ) (resi 119 and name HA ) 2.9 2.9 1.0
assi (resi 119 and name HD2# ) (resi 118 and name HN ) 5.5 5.5 1.0
assi (resi 119 and name HD2# ) (resi 120 and name HN ) 4.0 4.0 1.0
assi (resi 120 and name HN ) (resi 120 and name HB# ) 2.4 2.4 0.6
assi (resi 120 and name HA ) (resi 120 and name HB# ) 2.9 2.9 1.0
assi (resi 121 and name HN ) (resi 121 and name HB# ) 2.4 2.4 0.6
assi (resi 121 and name HN ) (resi 121 and name HB# ) 2.4 2.4 0.6
assi (resi 121 and name HN ) (resi 120 and name HB# ) 2.9 2.9 1.0
assi (resi 121 and name HB# ) (resi 121 and name HA ) 5.5 5.5 1.0
assi (resi 122 and name HN ) (resi 122 and name HB# ) 2.9 2.9 0.6
assi (resi 122 and name HN ) (resi 122 and name HB# ) 2.9 2.9 0.6
assi (resi 123 and name HA ) (resi 123 and name HB# ) 2.4 2.4 1.0
assi (resi 124 and name HN ) (resi 124 and name HB# ) 2.4 2.4 0.6
assi (resi 124 and name HN ) (resi 123 and name HB# ) 2.4 2.4 1.0
assi (resi 125 and name HN ) (resi 124 and name HB# ) 2.9 2.9 1.0
assi (resi 125 and name HN ) (resi 125 and name HG2# ) 4.0 4.0 1.5
assi (resi 125 and name HG2# ) (resi 125 and name HA ) 2.9 2.9 1.1
assi (resi 126 and name HN ) (resi 126 and name HB ) 2.9 2.9 0.0
assi (resi 126 and name HN ) (resi 126 and name HG1# ) 2.9 2.9 1.0
assi (resi 126 and name HA ) (resi 125 and name HG2# ) 4.0 4.0 1.5
assi (resi 126 and name HG2# ) (resi 126 and name HA ) 2.4 2.4 1.1
assi (resi 126 and name HG1# ) (resi 126 and name HA ) 3.4 3.4 0.6
assi (resi 126 and name HG1# ) (resi 126 and name HA ) 3.4 3.4 0.6
assi (resi 126 and name HD# ) (resi 126 and name HA ) 2.9 2.9 0.0
assi (resi 127 and name HN ) (resi 126 and name HB ) 2.4 2.4 0.0
assi (resi 127 and name HN ) (resi 126 and name HG1# ) 3.4 3.4 1.0
assi (resi 128 and name HN ) (resi 128 and name HE# ) 5.5 5.5 1.0
assi (resi 128 and name HN ) (resi 128 and name HB# ) 2.4 2.4 0.6
assi (resi 128 and name HN ) (resi 128 and name HD# ) 2.4 2.4 1.0
assi (resi 128 and name HN ) (resi 129 and name HD## ) 4.0 4.0 2.9
assi (resi 129 and name HN ) (resi 129 and name HB# ) 2.4 2.4 0.6
assi (resi 129 and name HN ) (resi 129 and name HD## ) 2.9 2.9 2.9
assi (resi 129 and name HN ) (resi 129 and name HD## ) 2.4 2.4 2.9
assi (resi 129 and name HA ) (resi 129 and name HB# ) 2.4 2.4 1.0
assi (resi 129 and name HA ) (resi 129 and name HD## ) 2.4 2.4 2.2
assi (resi 130 and name HN ) (resi 130 and name HB# ) 2.4 2.4 1.1
assi (resi 130 and name HN ) (resi 129 and name HD## ) 2.4 2.4 2.9
assi (resi 130 and name HA ) (resi 130 and name HB# ) 2.4 2.4 1.5
assi (resi 131 and name HN ) (resi 132 and name HB# ) 3.4 3.4 1.0
assi (resi 131 and name HN ) (resi 132 and name HB# ) 4.0 4.0 1.0
assi (resi 131 and name HN ) (resi 131 and name HB# ) 2.4 2.4 0.6
assi (resi 131 and name HN ) (resi 130 and name HB# ) 2.4 2.4 1.5
assi (resi 131 and name HN ) (resi 131 and name HG# ) 2.4 2.4 1.0
assi (resi 132 and name HN ) (resi 132 and name HB# ) 2.9 2.9 0.6
assi (resi 132 and name HN ) (resi 132 and name HB# ) 2.4 2.4 0.6
assi (resi 132 and name HN ) (resi 131 and name HB# ) 2.4 2.4 1.0
assi (resi 132 and name HN ) (resi 131 and name HG# ) 3.4 3.4 1.0
assi (resi 132 and name HD2# ) (resi 131 and name HA ) 4.0 4.0 0.0
assi (resi 132 and name HD2# ) (resi 132 and name HA ) 4.0 4.0 0.0
assi (resi 132 and name HD2# ) (resi 131 and name HA ) 4.0 4.0 0.0
assi (resi 132 and name HA ) (resi 131 and name HB# ) 3.4 3.4 1.0
assi (resi 132 and name HA ) (resi 131 and name HD# ) 4.0 4.0 1.0
assi (resi 132 and name HA ) (resi 133 and name HG# ) 3.4 3.4 1.0
assi (resi 132 and name HB# ) (resi 132 and name HA ) 3.4 3.4 1.0
assi (resi 132 and name HB# ) (resi 131 and name HA ) 5.5 5.5 1.0
assi (resi 132 and name HB# ) (resi 132 and name HA ) 3.4 3.4 1.0
assi (resi 132 and name HB# ) (resi 131 and name HA ) 5.5 5.5 1.0
assi (resi 133 and name HN ) (resi 133 and name HB# ) 2.9 2.9 0.6
assi (resi 133 and name HN ) (resi 133 and name HG# ) 2.9 2.9 1.0
assi (resi 133 and name HA ) (resi 133 and name HE# ) 5.5 5.5 1.0
assi (resi 133 and name HA ) (resi 133 and name HB# ) 2.9 2.9 1.0
assi (resi 133 and name HA ) (resi 133 and name HB# ) 2.9 2.9 1.0
assi (resi 133 and name HA ) (resi 133 and name HD# ) 2.9 2.9 1.0
assi (resi 133 and name HA ) (resi 133 and name HG# ) 2.9 2.9 0.6
assi (resi 134 and name HN ) (resi 133 and name HE# ) 5.5 5.5 1.0
assi (resi 134 and name HN ) (resi 134 and name HG# ) 2.9 2.9 1.0
assi (resi 134 and name HN ) (resi 134 and name HB# ) 2.9 2.9 0.6
assi (resi 134 and name HE2# ) (resi 134 and name HA ) 3.4 3.4 0.0
assi (resi 134 and name HE2# ) (resi 134 and name HA ) 3.4 3.4 0.0
assi (resi 134 and name HA ) (resi 134 and name HB# ) 4.0 4.0 1.0
assi (resi 135 and name HN ) (resi 134 and name HG# ) 4.0 4.0 1.0
assi (resi 135 and name HN ) (resi 134 and name HB# ) 2.4 2.4 1.0
assi (resi 135 and name HN ) (resi 135 and name HD# ) 2.4 2.4 1.0
assi (resi 135 and name HN ) (resi 135 and name HG# ) 2.4 2.4 1.0
assi (resi 137 and name HN ) (resi 137 and name HB ) 2.9 2.9 0.0
assi (resi 137 and name HN ) (resi 137 and name HG1# ) 3.4 3.4 1.0
assi (resi 137 and name HN ) (resi 137 and name HD# ) 2.9 2.9 0.0
assi (resi 137 and name HN ) (resi 137 and name HG2# ) 3.4 3.4 1.5
assi (resi 137 and name HD# ) (resi 137 and name HA ) 2.4 2.4 0.0
assi (resi 137 and name HG2# ) (resi 137 and name HA ) 2.4 2.4 1.1
assi (resi 138 and name HN ) (resi 138 and name HB ) 2.9 2.9 0.0
assi (resi 138 and name HN ) (resi 138 and name HG2# ) 2.4 2.4 1.5
assi (resi 138 and name HN ) (resi 137 and name HD# ) 5.5 5.5 0.0
assi (resi 138 and name HN ) (resi 137 and name HG2# ) 2.9 2.9 1.5
assi (resi 140 and name HN ) (resi 140 and name HG# ) 2.4 2.4 1.0
assi (resi 140 and name HN ) (resi 140 and name HB# ) 2.4 2.4 0.6
assi (resi 140 and name HA ) (resi 140 and name HG# ) 2.9 2.9 0.6
assi (resi 141 and name HN ) (resi 141 and name HB ) 2.9 2.9 0.0
assi (resi 141 and name HN ) (resi 140 and name HG# ) 3.4 3.4 1.0
assi (resi 141 and name HN ) (resi 140 and name HB# ) 2.4 2.4 1.0
assi (resi 141 and name HN ) (resi 141 and name HG2# ) 3.4 3.4 1.5
assi (resi 141 and name HN ) (resi 142 and name HB# ) 4.0 4.0 1.5
assi (resi 141 and name HA ) (resi 141 and name HB ) 2.9 2.9 0.0
assi (resi 141 and name HA ) (resi 141 and name HG2# ) 2.9 2.9 1.1
assi (resi 142 and name HN ) (resi 141 and name HB ) 3.4 3.4 0.0
assi (resi 142 and name HN ) (resi 141 and name HG2# ) 2.9 2.9 1.5
assi (resi 142 and name HN ) (resi 142 and name HB# ) 2.4 2.4 1.1
assi (resi 142 and name HA ) (resi 141 and name HG2# ) 3.4 3.4 1.5
assi (resi 142 and name HA ) (resi 142 and name HB# ) 2.9 2.9 1.5
assi (resi 143 and name HN ) (resi 143 and name HG# ) 2.9 2.9 1.0
assi (resi 143 and name HN ) (resi 143 and name HB# ) 2.4 2.4 0.6
assi (resi 143 and name HN ) (resi 142 and name HB# ) 2.9 2.9 1.5
assi (resi 143 and name HB# ) (resi 143 and name HA ) 3.4 3.4 1.0
assi (resi 143 and name HG# ) (resi 143 and name HA ) 2.9 2.9 0.6
assi (resi 143 and name HG# ) (resi 143 and name HA ) 2.9 2.9 0.6
assi (resi 144 and name HN ) (resi 144 and name HB# ) 2.4 2.4 0.6
assi (resi 145 and name HN ) (resi 144 and name HB# ) 2.4 2.4 1.0
assi (resi 145 and name HN ) (resi 145 and name HB# ) 2.4 2.4 0.6
assi (resi 145 and name HN ) (resi 145 and name HD1# ) 2.9 2.9 1.0
assi (resi 145 and name HN ) (resi 145 and name HD2# ) 3.4 3.4 1.0
assi (resi 145 and name HA ) (resi 145 and name HD1# ) 2.9 2.9 1.0
assi (resi 145 and name HA ) (resi 145 and name HD2# ) 4.0 4.0 1.0
assi (resi 146 and name HN ) (resi 145 and name HB# ) 2.9 2.9 1.0
assi (resi 146 and name HN ) (resi 146 and name HB# ) 2.4 2.4 1.1
assi (resi 146 and name HN ) (resi 145 and name HD1# ) 3.4 3.4 1.0
assi (resi 146 and name HN ) (resi 145 and name HD2# ) 4.0 4.0 1.0
assi (resi 146 and name HB# ) (resi 146 and name HA ) 2.4 2.4 1.5
assi (resi 147 and name HN ) (resi 147 and name HE ) 3.4 3.4 1.0
assi (resi 147 and name HN ) (resi 147 and name HB# ) 2.4 2.4 0.6
assi (resi 147 and name HN ) (resi 146 and name HB# ) 2.4 2.4 1.5
assi (resi 148 and name HN ) (resi 148 and name HB# ) 2.4 2.4 0.6
assi (resi 148 and name HN ) (resi 148 and name HB# ) 2.4 2.4 0.6
assi (resi 148 and name HN ) (resi 147 and name HB# ) 2.4 2.4 1.0
assi (resi 148 and name HA ) (resi 149 and name HD## ) 2.9 2.9 2.9
assi (resi 148 and name HB# ) (resi 148 and name HA ) 5.5 5.5 1.0
assi (resi 148 and name HB# ) (resi 148 and name HA ) 2.9 2.9 1.0
assi (resi 149 and name HN ) (resi 148 and name HB# ) 2.9 2.9 1.0
assi (resi 149 and name HN ) (resi 148 and name HB# ) 2.9 2.9 1.0
assi (resi 149 and name HN ) (resi 149 and name HB# ) 2.4 2.4 0.6
assi (resi 149 and name HD1# ) (resi 149 and name HN ) 3.4 3.4 1.0
assi (resi 149 and name HD1# ) (resi 149 and name HA ) 5.5 5.5 1.0
assi (resi 149 and name HD2# ) (resi 149 and name HN ) 2.9 2.9 1.0
assi (resi 149 and name HD2# ) (resi 149 and name HA ) 2.9 2.9 1.0
assi (resi 150 and name HN ) (resi 150 and name HB# ) 3.4 3.4 0.6
assi (resi 150 and name HN ) (resi 150 and name HD# ) 2.4 2.4 1.0
assi (resi 150 and name HN ) (resi 150 and name HG# ) 2.4 2.4 1.0
assi (resi 150 and name HN ) (resi 149 and name HD## ) 3.4 3.4 2.9
assi (resi 150 and name HA ) (resi 150 and name HB# ) 2.9 2.9 1.0
assi (resi 150 and name HA ) (resi 150 and name HD# ) 2.4 2.4 1.0
assi (resi 150 and name HA ) (resi 150 and name HG# ) 2.4 2.4 0.6
assi (resi 151 and name HN ) (resi 150 and name HD# ) 4.0 4.0 1.0
assi (resi 151 and name HN ) (resi 151 and name HB# ) 2.4 2.4 1.1
assi (resi 151 and name HA ) (resi 151 and name HB# ) 2.9 2.9 1.5
assi (resi 151 and name HB# ) (resi 152 and name HN ) 3.4 3.4 1.5
assi (resi 152 and name HN ) (resi 152 and name HB ) 2.9 2.9 0.0
assi (resi 152 and name HA ) (resi 152 and name HB ) 3.4 3.4 0.0
assi (resi 152 and name HB ) (resi 153 and name HN ) 5.5 5.5 0.0
assi (resi 152 and name HG2# ) (resi 152 and name HN ) 2.9 2.9 1.5
assi (resi 152 and name HG2# ) (resi 153 and name HN ) 2.9 2.9 1.5
assi (resi 152 and name HG2# ) (resi 152 and name HA ) 2.4 2.4 1.1
assi (resi 152 and name HG1# ) (resi 152 and name HN ) 4.0 4.0 1.5
assi (resi 152 and name HG1# ) (resi 153 and name HN ) 2.9 2.9 1.5
assi (resi 152 and name HG1# ) (resi 152 and name HA ) 2.4 2.4 1.1
assi (resi 153 and name HN ) (resi 153 and name HB# ) 2.4 2.4 0.6
assi (resi 154 and name HN ) (resi 154 and name HB# ) 2.4 2.4 0.6
assi (resi 154 and name HN ) (resi 153 and name HB# ) 2.9 2.9 1.0
assi (resi 154 and name HA ) (resi 154 and name HD# ) 5.5 5.5 1.0
assi (resi 154 and name HA ) (resi 154 and name HB# ) 5.5 5.5 1.0
assi (resi 155 and name HN ) (resi 154 and name HB# ) 3.4 3.4 1.0
assi (resi 155 and name HA ) (resi 155 and name HB ) 2.9 2.9 0.0
assi (resi 155 and name HG1# ) (resi 155 and name HN ) 3.4 3.4 1.5
assi (resi 155 and name HG1# ) (resi 154 and name HA ) 4.0 4.0 1.5
assi (resi 155 and name HG1# ) (resi 155 and name HA ) 2.9 2.9 1.1
assi (resi 155 and name HG2# ) (resi 155 and name HN ) 3.4 3.4 1.5
assi (resi 155 and name HG2# ) (resi 154 and name HA ) 4.0 4.0 1.5
assi (resi 155 and name HG2# ) (resi 155 and name HA ) 3.4 3.4 1.1
assi (resi 156 and name HN ) (resi 156 and name HG## ) 2.4 2.4 1.5
assi (resi 156 and name HN ) (resi 155 and name HB ) 2.4 2.4 0.0
assi (resi 156 and name HN ) (resi 155 and name HG## ) 2.4 2.4 2.9
assi (resi 156 and name HA ) (resi 156 and name HG# ) 5.5 5.5 0.6
assi (resi 157 and name HA ) (resi 157 and name HB# ) 5.5 5.5 1.0
assi (resi 157 and name HA ) (resi 157 and name HB# ) 5.5 5.5 1.0
assi (resi 158 and name HN ) (resi 157 and name HB# ) 3.4 3.4 1.0
assi (resi 158 and name HN ) (resi 157 and name HB# ) 3.4 3.4 1.0
assi (resi 159 and name HN ) (resi 159 and name HB# ) 2.9 2.9 1.1
assi (resi 159 and name HN ) (resi 160 and name HB# ) 5.5 5.5 1.0
assi (resi 159 and name HA ) (resi 159 and name HB# ) 2.9 2.9 1.5
assi (resi 159 and name HB# ) (resi 160 and name HN ) 4.0 4.0 1.5
assi (resi 159 and name HB# ) (resi 158 and name HA ) 5.5 5.5 1.5
assi (resi 160 and name HN ) (resi 160 and name HB# ) 2.9 2.9 0.6
assi (resi 160 and name HN ) (resi 160 and name HD1# ) 3.4 3.4 1.0
assi (resi 160 and name HA ) (resi 160 and name HB# ) 4.0 4.0 1.0
assi (resi 160 and name HG ) (resi 160 and name HN ) 5.5 5.5 0.0
assi (resi 160 and name HD2# ) (resi 160 and name HN ) 3.4 3.4 1.0
assi (resi 160 and name HD2# ) (resi 161 and name HA ) 4.0 4.0 1.0
assi (resi 160 and name HD1# ) (resi 160 and name HN ) 3.4 3.4 1.0
assi (resi 160 and name HD1# ) (resi 161 and name HA ) 4.0 4.0 1.0
assi (resi 161 and name HN ) (resi 162 and name HG# ) 5.5 5.5 1.0
assi (resi 161 and name HN ) (resi 161 and name HG2# ) 2.4 2.4 1.5
assi (resi 161 and name HN ) (resi 160 and name HG ) 3.4 3.4 0.0
assi (resi 161 and name HN ) (resi 160 and name HD2# ) 3.4 3.4 1.0
assi (resi 161 and name HN ) (resi 160 and name HD1# ) 3.4 3.4 1.0
assi (resi 161 and name HA ) (resi 161 and name HG2# ) 2.4 2.4 1.1
assi (resi 162 and name HN ) (resi 162 and name HB# ) 3.4 3.4 0.6
assi (resi 162 and name HN ) (resi 162 and name HG# ) 2.9 2.9 1.0
assi (resi 162 and name HN ) (resi 161 and name HG2# ) 2.9 2.9 1.5
assi (resi 162 and name HE2# ) (resi 162 and name HA ) 4.0 4.0 0.0
assi (resi 162 and name HE2# ) (resi 162 and name HA ) 3.4 3.4 0.0
assi (resi 162 and name HA ) (resi 161 and name HG2# ) 3.4 3.4 1.5
assi (resi 162 and name HB# ) (resi 162 and name HA ) 4.0 4.0 1.0
assi (resi 162 and name HB# ) (resi 162 and name HA ) 4.0 4.0 1.0
assi (resi 162 and name HG# ) (resi 162 and name HA ) 2.9 2.9 0.6
assi (resi 162 and name HG# ) (resi 162 and name HA ) 3.4 3.4 0.6
assi (resi 163 and name HN ) (resi 162 and name HB# ) 5.5 5.5 1.0
assi (resi 163 and name HN ) (resi 162 and name HG# ) 4.0 4.0 1.0
assi (resi 163 and name HN ) (resi 163 and name HB# ) 2.4 2.4 0.6
assi (resi 163 and name HN ) (resi 163 and name HG# ) 2.4 2.4 1.0
assi (resi 164 and name HN ) (resi 163 and name HB# ) 2.9 2.9 1.0
assi (resi 164 and name HN ) (resi 163 and name HG# ) 2.4 2.4 1.0
assi (resi 165 and name HN ) (resi 165 and name HB# ) 2.4 2.4 0.6
assi (resi 165 and name HN ) (resi 165 and name HG ) 3.4 3.4 0.0
assi (resi 165 and name HN ) (resi 165 and name HD## ) 2.4 2.4 2.9
assi (resi 165 and name HA ) (resi 165 and name HB# ) 3.4 3.4 1.0
assi (resi 165 and name HD1# ) (resi 165 and name HA ) 2.9 2.9 1.0
assi (resi 166 and name HN ) (resi 167 and name HB# ) 3.4 3.4 1.0
assi (resi 166 and name HN ) (resi 166 and name HB# ) 2.4 2.4 0.6
assi (resi 166 and name HN ) (resi 165 and name HD## ) 2.4 2.4 2.9
assi (resi 166 and name HA ) (resi 166 and name HB# ) 3.4 3.4 1.0
assi (resi 167 and name HN ) (resi 167 and name HB# ) 2.4 2.4 0.6
assi (resi 167 and name HN ) (resi 167 and name HB# ) 2.4 2.4 0.6
assi (resi 167 and name HN ) (resi 166 and name HB# ) 2.4 2.4 1.0
assi (resi 167 and name HN ) (resi 168 and name HG2# ) 4.0 4.0 1.5
assi (resi 167 and name HN ) (resi 168 and name HG1# ) 5.5 5.5 1.5
assi (resi 167 and name HD2# ) (resi 168 and name HB ) 2.9 2.9 0.0
assi (resi 167 and name HB# ) (resi 167 and name HA ) 2.9 2.9 1.0
assi (resi 167 and name HB# ) (resi 167 and name HA ) 2.9 2.9 1.0
assi (resi 168 and name HN ) (resi 167 and name HB# ) 2.4 2.4 1.0
assi (resi 168 and name HN ) (resi 168 and name HB ) 2.4 2.4 0.0
assi (resi 168 and name HN ) (resi 168 and name HG2# ) 2.4 2.4 1.5
assi (resi 168 and name HN ) (resi 168 and name HG1# ) 2.9 2.9 1.5
assi (resi 168 and name HA ) (resi 168 and name HB ) 3.4 3.4 0.0
assi (resi 168 and name HG2# ) (resi 168 and name HA ) 2.4 2.4 1.1
assi (resi 168 and name HG1# ) (resi 169 and name HA ) 3.4 3.4 1.5
assi (resi 168 and name HG1# ) (resi 168 and name HA ) 2.9 2.9 1.1
assi (resi 169 and name HN ) (resi 169 and name HB# ) 2.9 2.9 0.6
assi (resi 169 and name HN ) (resi 168 and name HB ) 2.9 2.9 0.0
assi (resi 169 and name HN ) (resi 168 and name HG2# ) 2.9 2.9 1.5
assi (resi 169 and name HN ) (resi 168 and name HG1# ) 2.9 2.9 1.5
assi (resi 170 and name HN ) (resi 170 and name HB# ) 2.9 2.9 0.6
assi (resi 172 and name HN ) (resi 172 and name HB# ) 2.4 2.4 1.1
assi (resi 172 and name HB# ) (resi 172 and name HA ) 2.4 2.4 1.5
assi (resi 173 and name HN ) (resi 172 and name HB# ) 2.4 2.4 1.5
assi (resi 173 and name HA ) (resi 173 and name HB ) 3.4 3.4 0.0
assi (resi 173 and name HB ) (resi 173 and name HN ) 3.4 3.4 0.0
assi (resi 173 and name HG2# ) (resi 173 and name HN ) 2.9 2.9 1.5
assi (resi 173 and name HG2# ) (resi 174 and name HA ) 3.4 3.4 1.5
assi (resi 173 and name HG2# ) (resi 173 and name HA ) 3.4 3.4 1.1
assi (resi 173 and name HD# ) (resi 173 and name HN ) 2.9 2.9 0.0
assi (resi 173 and name HD# ) (resi 173 and name HA ) 3.4 3.4 0.0
assi (resi 174 and name HB# ) (resi 174 and name HA ) 4.0 4.0 1.0
assi (resi 174 and name HB# ) (resi 174 and name HN ) 4.0 4.0 0.6
assi (resi 174 and name HB# ) (resi 174 and name HA ) 4.0 4.0 1.0
assi (resi 174 and name HD1# ) (resi 174 and name HN ) 2.9 2.9 1.0
assi (resi 174 and name HD1# ) (resi 174 and name HA ) 2.9 2.9 1.0
assi (resi 174 and name HD2# ) (resi 174 and name HA ) 2.4 2.4 1.0
assi (resi 175 and name HN ) (resi 174 and name HD1# ) 2.4 2.4 1.0
assi (resi 175 and name HA ) (resi 175 and name HB# ) 2.4 2.4 1.5
assi (resi 175 and name HB# ) (resi 175 and name HN ) 2.9 2.9 1.1
assi (resi 176 and name HN ) (resi 177 and name HB# ) 3.4 3.4 1.0
assi (resi 176 and name HB# ) (resi 176 and name HN ) 2.9 2.9 1.1
assi (resi 176 and name HB# ) (resi 176 and name HA ) 2.9 2.9 1.5
assi (resi 177 and name HN ) (resi 176 and name HB# ) 2.4 2.4 1.5
assi (resi 177 and name HN ) (resi 177 and name HD2# ) 3.4 3.4 1.0
assi (resi 177 and name HN ) (resi 177 and name HD1# ) 2.9 2.9 1.0
assi (resi 177 and name HB# ) (resi 177 and name HA ) 3.4 3.4 1.0
assi (resi 177 and name HB# ) (resi 177 and name HA ) 4.0 4.0 1.0
assi (resi 177 and name HG ) (resi 177 and name HA ) 4.0 4.0 0.0
assi (resi 177 and name HD2# ) (resi 176 and name HN ) 5.5 5.5 1.0
assi (resi 177 and name HD2# ) (resi 177 and name HA ) 2.4 2.4 1.0
assi (resi 177 and name HD1# ) (resi 176 and name HN ) 5.5 5.5 1.0
assi (resi 177 and name HD1# ) (resi 177 and name HA ) 2.9 2.9 1.0
assi (resi 178 and name HN ) (resi 179 and name HD# ) 2.9 2.9 1.0
assi (resi 178 and name HN ) (resi 178 and name HG# ) 3.4 3.4 1.0
assi (resi 178 and name HN ) (resi 178 and name HB# ) 2.4 2.4 0.6
assi (resi 178 and name HN ) (resi 177 and name HB# ) 2.9 2.9 1.0
assi (resi 178 and name HN ) (resi 177 and name HB# ) 2.9 2.9 1.0
assi (resi 178 and name HN ) (resi 177 and name HD2# ) 3.4 3.4 1.0
assi (resi 178 and name HN ) (resi 177 and name HD1# ) 3.4 3.4 1.0
assi (resi 178 and name HA ) (resi 179 and name HD# ) 2.4 2.4 1.0
assi (resi 179 and name HB# ) (resi 179 and name HD# ) 2.9 2.9 1.0
assi (resi 179 and name HB# ) (resi 179 and name HA ) 4.0 4.0 1.0
assi (resi 180 and name HA ) (resi 180 and name HB# ) 2.4 2.4 1.0
assi (resi 180 and name HA ) (resi 180 and name HG# ) 2.9 2.9 0.6
assi (resi 180 and name HA ) (resi 180 and name HB# ) 2.9 2.9 1.0
assi (resi 181 and name HN ) (resi 182 and name HD# ) 2.9 2.9 1.0
assi (resi 181 and name HN ) (resi 181 and name HG# ) 2.4 2.4 1.0
assi (resi 181 and name HN ) (resi 181 and name HB# ) 2.4 2.4 0.6
assi (resi 182 and name HA ) (resi 182 and name HD# ) 4.0 4.0 1.0
assi (resi 182 and name HA ) (resi 182 and name HB# ) 2.4 2.4 1.0
assi (resi 182 and name HA ) (resi 182 and name HG# ) 2.9 2.9 0.6
assi (resi 182 and name HA ) (resi 182 and name HB# ) 2.4 2.4 1.0
assi (resi 183 and name HN ) (resi 182 and name HD# ) 3.4 3.4 1.0
assi (resi 183 and name HN ) (resi 183 and name HE# ) 5.5 5.5 1.0
assi (resi 183 and name HN ) (resi 182 and name HB# ) 3.4 3.4 1.0
assi (resi 183 and name HN ) (resi 183 and name HB# ) 2.4 2.4 0.6
assi (resi 183 and name HN ) (resi 183 and name HG# ) 2.9 2.9 1.0
assi (resi 185 and name HN ) (resi 185 and name HB# ) 2.9 2.9 0.6
assi (resi 186 and name HN ) (resi 186 and name HB# ) 5.5 5.5 0.6
assi (resi 186 and name HN ) (resi 186 and name HG# ) 3.4 3.4 1.0
assi (resi 187 and name HN ) (resi 187 and name HD# ) 5.5 5.5 1.0
assi (resi 187 and name HN ) (resi 187 and name HB# ) 2.4 2.4 0.6

!INTRARESIDUE AND SEQUENTIAL SIDECHAIN TO SIDECHAIN

assi (resi  -4 and name HB ) (resi  -4 and name HG1# ) 4.0 4.0 1.0
assi (resi  -4 and name HB ) (resi  -4 and name HG1# ) 4.0 4.0 1.0
assi (resi  -4 and name HB ) (resi  -4 and name HG2# ) 2.4 2.4 1.5
assi (resi  -4 and name HB ) (resi  -4 and name HD# ) 2.4 2.4 0.0
assi (resi  -3 and name HB ) (resi  -3 and name HG1# ) 3.4 3.4 1.0
assi (resi  -3 and name HB ) (resi  -3 and name HG1# ) 3.4 3.4 1.0
assi (resi  -3 and name HB ) (resi  -3 and name HG2# ) 2.4 2.4 1.5
assi (resi  -3 and name HB ) (resi  -3 and name HD# ) 2.4 2.4 0.0
assi (resi   1 and name HB# ) (resi   1 and name HG# ) 2.4 2.4 1.0
assi (resi   1 and name HB# ) (resi   1 and name HG# ) 2.9 2.9 1.0
assi (resi   2 and name HE2# ) (resi   2 and name HE2# ) 2.4 2.4 0.0
assi (resi   2 and name HE2# ) (resi   2 and name HG# ) 2.4 2.4 1.0
assi (resi   2 and name HE2# ) (resi   2 and name HB# ) 2.9 2.9 1.0
assi (resi   2 and name HE2# ) (resi   2 and name HG# ) 2.9 2.9 1.0
assi (resi   4 and name HG2# ) (resi   4 and name HB ) 2.9 2.9 1.5
assi (resi   4 and name HD# ) (resi   4 and name HB ) 2.4 2.4 0.0
assi (resi   4 and name HD# ) (resi   4 and name HG1# ) 2.9 2.9 1.0
assi (resi   7 and name HB ) (resi   7 and name HG2# ) 2.9 2.9 1.5
assi (resi   7 and name HB ) (resi   7 and name HG1# ) 2.4 2.4 1.5
assi (resi   8 and name HG2# ) (resi   8 and name HB ) 2.9 2.9 1.5
assi (resi   8 and name HG1# ) (resi   8 and name HB ) 2.9 2.9 1.5
assi (resi   9 and name HG1# ) (resi   9 and name HG2# ) 2.4 2.4 2.0
assi (resi   9 and name HG2# ) (resi   9 and name HB ) 2.4 2.4 1.5
assi (resi  19 and name HB# ) (resi  19 and name HB# ) 2.9 2.9 1.0
assi (resi  19 and name HB# ) (resi  19 and name HB# ) 2.9 2.9 1.0
assi (resi  19 and name HG ) (resi  19 and name HD1# ) 2.4 2.4 1.0
assi (resi  19 and name HG ) (resi  19 and name HD2# ) 2.4 2.4 1.0
assi (resi  19 and name HD1# ) (resi  19 and name HB# ) 3.4 3.4 1.0
assi (resi  19 and name HD1# ) (resi  19 and name HB# ) 2.4 2.4 1.0
assi (resi  19 and name HD2# ) (resi  19 and name HB# ) 2.9 2.9 1.0
assi (resi  19 and name HD2# ) (resi  19 and name HB# ) 2.4 2.4 1.0
assi (resi  22 and name HB# ) (resi  21 and name HG2# ) 3.4 3.4 2.5
assi (resi  24 and name HB ) (resi  24 and name HG2# ) 2.4 2.4 1.5
assi (resi  25 and name HB ) (resi  25 and name HG2# ) 2.4 2.4 1.5
assi (resi  25 and name HG2# ) (resi  24 and name HB ) 3.4 3.4 1.5
assi (resi  26 and name HD2# ) (resi  26 and name HD2# ) 2.4 2.4 0.0
assi (resi  26 and name HD2# ) (resi  26 and name HB# ) 2.4 2.4 1.0
assi (resi  26 and name HD2# ) (resi  26 and name HB# ) 2.9 2.9 1.0
assi (resi  26 and name HD2# ) (resi  26 and name HB# ) 2.9 2.9 1.0
assi (resi  26 and name HB# ) (resi  26 and name HB# ) 2.9 2.9 1.0
assi (resi  26 and name HB# ) (resi  27 and name HB# ) 4.0 4.0 2.0
assi (resi  26 and name HB# ) (resi  27 and name HG# ) 5.5 5.5 2.0
assi (resi  26 and name HB# ) (resi  27 and name HB# ) 5.5 5.5 2.0
assi (resi  26 and name HB# ) (resi  27 and name HG# ) 5.5 5.5 2.0
assi (resi  27 and name HB# ) (resi  27 and name HE# ) 4.0 4.0 1.0
assi (resi  27 and name HB# ) (resi  27 and name HG# ) 2.4 2.4 1.0
assi (resi  27 and name HB# ) (resi  27 and name HE# ) 4.0 4.0 1.0
assi (resi  27 and name HB# ) (resi  27 and name HG# ) 2.4 2.4 1.0
assi (resi  28 and name HB# ) (resi  28 and name HD# ) 5.5 5.5 1.0
assi (resi  28 and name HB# ) (resi  28 and name HD# ) 5.5 5.5 1.0
```

```
assi (resi  31 and name HG# ) (resi  30 and name HB# ) 3.4 3.4 2.0
assi (resi  32 and name HB# ) (resi  32 and name HD# ) 2.9 2.9 1.0
assi (resi  32 and name HB# ) (resi  32 and name HB# ) 2.4 2.4 1.0
assi (resi  32 and name HB# ) (resi  32 and name HD# ) 2.9 2.9 1.0
assi (resi  33 and name HB ) (resi  34 and name HD# ) 5.5 5.5 1.0
assi (resi  33 and name HB ) (resi  33 and name HG2# ) 2.4 2.4 1.5
assi (resi  33 and name HB ) (resi  33 and name HG1# ) 2.4 2.4 1.5
assi (resi  39 and name HD2# ) (resi  39 and name HD2# ) 2.4 2.4 0.0
assi (resi  42 and name HG1# ) (resi  42 and name HB ) 2.4 2.4 1.5
assi (resi  42 and name HG2# ) (resi  42 and name HB ) 2.4 2.4 1.5
assi (resi  43 and name HB ) (resi  43 and name HG2# ) 2.4 2.4 1.5
assi (resi  44 and name HG1# ) (resi  44 and name HB ) 2.4 2.4 1.5
assi (resi  44 and name HG2# ) (resi  44 and name HB ) 2.4 2.4 1.5
assi (resi  45 and name HE# ) (resi  45 and name HG# ) 2.9 2.9 1.5
assi (resi  46 and name HG2# ) (resi  46 and name HB ) 2.4 2.4 1.5
assi (resi  46 and name HG2# ) (resi 46 and name HG1# ) 2.9 2.9 1.5
assi (resi  46 and name HD# ) (resi  46 and name HB ) 2.4 2.4 0.0
assi (resi  46 and name HD# ) (resi  46 and name HG1# ) 2.9 2.9 1.0
assi (resi  52 and name HB ) (resi  52 and name HG2# ) 2.4 2.4 1.5
assi (resi  53 and name HD2# ) (resi  53 and name HB# ) 2.9 2.9 1.0
assi (resi  61 and name HE2# ) (resi  61 and name HE2# ) 2.4 2.4 0.0
assi (resi  61 and name HE2# ) (resi  61 and name HG# ) 2.4 2.4 1.0
assi (resi  61 and name HE2# ) (resi  61 and name HG# ) 2.9 2.9 1.0
assi (resi  62 and name HG# ) (resi  62 and name HB# ) 2.4 2.4 1.0
assi (resi  63 and name HB# ) (resi  63 and name HB# ) 2.4 2.4 1.0
assi (resi  64 and name HB# ) (resi  64 and name HB# ) 4.0 4.0 1.0
assi (resi  64 and name HB# ) (resi  64 and name HD# ) 4.0 4.0 1.0
assi (resi  70 and name HB# ) (resi  70 and name HB# ) 3.4 3.4 1.0
assi (resi  70 and name HB# ) (resi  70 and name HD1# ) 2.9 2.9 1.0
assi (resi  70 and name HB# ) (resi  70 and name HD2# ) 3.4 3.4 1.0
assi (resi  70 and name HB# ) (resi  70 and name HD1# ) 3.4 3.4 1.0
assi (resi  70 and name HB# ) (resi  70 and name HD2# ) 2.9 2.9 1.0
assi (resi  74 and name HE2# ) (resi  74 and name HE2# ) 2.4 2.4 0.0
assi (resi  74 and name HE2# ) (resi  74 and name HG# ) 3.4 3.4 1.0
assi (resi  74 and name HE2# ) (resi  74 and name HG# ) 3.4 3.4 1.0
assi (resi  77 and name HB ) (resi  78 and name HD# ) 2.9 2.9 2.0
assi (resi  77 and name HB ) (resi  78 and name HE# ) 4.0 4.0 2.0
assi (resi  77 and name HG1# ) (resi 77 and name HG2# ) 4.0 4.0 2.0
assi (resi  77 and name HG2# ) (resi  77 and name HB ) 2.9 2.9 1.5
assi (resi  79 and name HD1# ) (resi  79 and name HG ) 2.9 2.9 1.0
assi (resi  79 and name HD2# ) (resi  79 and name HD1# ) 2.9 2.9 1.0
assi (resi  80 and name HB ) (resi  80 and name HG2# ) 2.9 2.9 1.5
assi (resi  80 and name HB ) (resi  80 and name HG1# ) 2.9 2.9 1.5
assi (resi  80 and name HG2# ) (resi 80 and name HG1# ) 2.4 2.4 2.0
assi (resi  80 and name HG1# ) (resi  80 and name HG2# ) 2.9 2.9 2.
assi (resi  84 and name HG1# ) (resi  84 and name HB ) 2.9 2.9 1.5
assi (resi  84 and name HG1# ) (resi 85 and name HG1# ) 4.0 4.0 3.0
assi (resi  85 and name HG1# ) (resi  85 and name HB ) 4.0 4.0 1.5
assi (resi  85 and name HG2# ) (resi  85 and name HB ) 2.9 2.9 1.5
assi (resi  88 and name HB# ) (resi  88 and name HB# ) 2.9 2.9 1.0
assi (resi  92 and name HD2# ) (resi  92 and name HB# ) 2.9 2.9 1.0
assi (resi  92 and name HD2# ) (resi  92 and name HB# ) 2.9 2.9 1.0
assi (resi  92 and name HD2# ) (resi  92 and name HD2# ) 2.4 2.4 0.0
assi (resi  92 and name HD2# ) (resi  92 and name HB# ) 2.9 2.9 1.0
assi (resi  92 and name HD2# ) (resi  92 and name HB# ) 2.9 2.9 1.0
assi (resi  92 and name HB# ) (resi  92 and name HB# ) 3.4 3.4 1.0
assi (resi  94 and name HB# ) (resi  94 and name HE# ) 5.5 5.5 1.0
assi (resi  94 and name HB# ) (resi  94 and name HG# ) 2.9 2.9 1.0
assi (resi  94 and name HB# ) (resi  94 and name HE# ) 4.0 4.0 1.0
assi (resi  94 and name HB# ) (resi  94 and name HG# ) 3.4 3.4 1.0
assi (resi  94 and name HG# ) (resi  94 and name HE# ) 2.4 2.4 1.0
assi (resi  94 and name HD# ) (resi  94 and name HE# ) 2.4 2.4 1.0
assi (resi  94 and name HD# ) (resi  94 and name HG# ) 2.4 2.4 1.0
assi (resi  95 and name HB# ) (resi  95 and name HG# ) 2.4 2.4 1.0
assi (resi  97 and name HE1 ) (resi  97 and name HZ2 ) 3.4 3.4 0.0
assi (resi  97 and name HE1 ) (resi  97 and name HD1 ) 2.9 2.9 0.0
assi (resi  98 and name HG1# ) (resi  97 and name HB# ) 3.4 3.4 2.5
assi (resi  98 and name HG2# ) (resi  97 and name HB# ) 5.5 5.5 2.5
assi (resi  98 and name HG2# ) (resi  98 and name HB ) 2.9 2.9 1.5
assi (resi 101 and name HG2# ) (resi 100 and name HB# ) 3.4 3.4 2.5
assi (resi 101 and name HG2#)(resi 102 and name HG2#) 4.0 4.0 3.0
assi (resi 101 and name HG2# ) (resi 101 and name HD#) 3.4 3.4 1.5
assi (resi 101 and name HD# ) (resi 102 and name HG2#) 3.4 3.4 1.5
assi (resi 101 and name HD# ) (resi 101 and name HG1#) 2.4 2.4 1.0
assi (resi 102 and name HA ) (resi 101 and name HB ) 4.0 4.0 0.0
assi (resi 102 and name HB ) (resi 102 and name HG2# ) 2.4 2.4 1.5
assi (resi 111 and name HB# ) (resi 111 and name HD## ) 2.9 2.9 1.5
assi (resi 111 and name HB# ) (resi 111 and name HD## ) 2.9 2.9 1.5
assi (resi 111 and name HG ) (resi 111 and name HD## ) 2.4 2.4 2.9
assi (resi 112 and name HB# ) (resi 112 and name HB# ) 3.4 3.4 1.0
assi (resi 112 and name HD1# ) (resi 112 and name HG ) 2.9 2.9 1.0
assi (resi 112 and name HD2# ) (resi 112 and name HG ) 2.9 2.9 1.0
assi (resi 113 and name HG1# ) (resi 113 and name HB ) 2.9 2.9 1.5
assi (resi 116 and name HE2#) (resi 116 and name HE2#) 2.4 2.4 0.0
assi (resi 117 and name HB ) (resi 117 and name HG2# ) 3.4 3.4 1.5
assi (resi 117 and name HB ) (resi 117 and name HD# ) 3.4 3.4 0.0
assi (resi 119 and name HD1# ) (resi 119 and name HB# ) 2.4 2.4 1.0
assi (resi 126 and name HB ) (resi 126 and name HG1# ) 3.4 3.4 1.0
assi (resi 126 and name HB ) (resi 126 and name HG1# ) 5.5 5.5 1.0
assi (resi 126 and name HG2# ) (resi 126 and name HB ) 2.4 2.4 1.5
assi (resi 126 and name HG1#)(resi 126 and name HG1#) 2.9 2.9 1.0
assi (resi 126 and name HD# ) (resi 126 and name HG1#) 2.4 2.4 1.0
assi (resi 126 and name HD# ) (resi 126 and name HG1#) 3.4 3.4 1.0
assi (resi 132 and name HD2# ) (resi 132 and name HB# ) 2.4 2.4 1.0
assi (resi 132 and name HD2# ) (resi 132 and name HB# ) 2.9 2.9 1.0
assi (resi 132 and name HD2# ) (resi 131 and name HB# ) 4.0 4.0 1.0
assi (resi 132 and name HD2# ) (resi 131 and name HG#) 4.0 4.0 1.0
assi (resi 132 and name HD2#) (resi 132 and name HD2#) 2.4 2.4 0.0
assi (resi 132 and name HD2# ) (resi 132 and name HB# ) 2.9 2.9 1.0
assi (resi 132 and name HD2# ) (resi 132 and name HB# ) 2.9 2.9 1.0
assi (resi 132 and name HD2# ) (resi 131 and name HB# ) 3.4 3.4 1.0
assi (resi 132 and name HD2# ) (resi 131 and name HG#) 4.0 4.0 1.0
assi (resi 132 and name HB# ) (resi 132 and name HB# ) 2.4 2.4 1.0
assi (resi 132 and name HB# ) (resi 131 and name HB# ) 4.0 4.0 2.0
assi (resi 134 and name HE2# ) (resi 134 and name HG# ) 2.4 2.4 1.0
assi (resi 134 and name HE2# ) (resi 134 and name HB# ) 2.4 2.4 1.0
assi (resi 134 and name HE2#) (resi 134 and name HE2#) 2.4 2.4 0.0
assi (resi 134 and name HE2# ) (resi 134 and name HG# ) 2.4 2.4 1.0
assi (resi 134 and name HE2# ) (resi 134 and name HB# ) 2.9 2.9 1.0
assi (resi 137 and name HB ) (resi 137 and name HG2# ) 3.4 3.4 1.5
assi (resi 137 and name HD# ) (resi 137 and name HB ) 2.9 2.9 0.0
assi (resi 141 and name HB ) (resi 141 and name HG2# ) 2.4 2.4 1.5
assi (resi 141 and name HB ) (resi 142 and name HB# ) 2.9 2.9 1.5
assi (resi 141 and name HG2# ) (resi 142 and name HB# ) 2.4 2.4 3.0
assi (resi 143 and name HG# ) (resi 143 and name HB# ) 2.9 2.9 1.0
assi (resi 143 and name HG# ) (resi 142 and name HB# ) 3.4 3.4 2.5
assi (resi 143 and name HG# ) (resi 143 and name HB# ) 2.4 2.4 1.0
assi (resi 143 and name HG# ) (resi 142 and name HB# ) 4.0 4.0 2.5
assi (resi 145 and name HG ) (resi 145 and name HD1# ) 2.9 2.9 1.0
assi (resi 145 and name HG ) (resi 145 and name HD2# ) 3.4 3.4 1.0
assi (resi 145 and name HD1#) (resi 145 and name HD2#) 2.4 2.4 1.0
assi (resi 146 and name HB# ) (resi 145 and name HD2# ) 5.5 5.5 2.5
assi (resi 147 and name HE ) (resi 147 and name HB# ) 3.4 3.4 1.0
assi (resi 147 and name HE ) (resi 147 and name HG# ) 5.5 5.5 1.0
assi (resi 149 and name HD1# ) (resi 149 and name HB# ) 2.4 2.4 1.0
assi (resi 149 and name HD2# ) (resi 149 and name HB# ) 2.4 2.4 1.0
assi (resi 152 and name HB ) (resi 152 and name HG2# ) 2.9 2.9 1.5
assi (resi 152 and name HB ) (resi 152 and name HG1# ) 2.9 2.9 1.5
assi (resi 155 and name HG1# ) (resi 155 and name HB ) 2.4 2.4 1.5
assi (resi 155 and name HG2# ) (resi 155 and name HB ) 2.4 2.4 1.5
assi (resi 159 and name HB# ) (resi 160 and name HD1# ) 4.0 4.0 2.5
assi (resi 160 and name HB# ) (resi 160 and name HD2# ) 3.4 3.4 1.0
assi (resi 160 and name HB# ) (resi 160 and name HD1# ) 3.4 3.4 1.0
assi (resi 160 and name HB# ) (resi 160 and name HB# ) 4.0 4.0 1.0
assi (resi 160 and name HB# ) (resi 160 and name HD2# ) 3.4 3.4 1.0
assi (resi 160 and name HB# ) (resi 160 and name HD1# ) 3.4 3.4 1.0
assi (resi 160 and name HG ) (resi 160 and name HD2# ) 2.9 2.9 1.0
assi (resi 160 and name HG ) (resi 160 and name HD1# ) 2.4 2.4 1.0
assi (resi 161 and name HB ) (resi 161 and name HG2# ) 2.9 2.9 1.5
assi (resi 161 and name HG2#)(resi 160 and name HD2#) 2.9 2.9 2.5
assi (resi 162 and name HE2# ) (resi 161 and name HB ) 3.4 3.4 0.0
assi (resi 162 and name HE2# ) (resi 162 and name HG# ) 2.4 2.4 1.0
assi (resi 162 and name HE2#) (resi 161 and name HG2#) 2.9 2.9 1.5
assi (resi 162 and name HE2#) (resi 162 and name HE2#) 2.4 2.4 0.0
assi (resi 162 and name HE2# ) (resi 161 and name HB ) 3.4 3.4 0.0
assi (resi 162 and name HE2# ) (resi 162 and name HG# ) 2.9 2.9 1.0
assi (resi 162 and name HE2#) (resi 161 and name HG2#) 2.9 2.9 1.5
assi (resi 162 and name HB# ) (resi 161 and name HG2# ) 4.0 4.0 2.5
assi (resi 162 and name HB# ) (resi 162 and name HB# ) 3.4 3.4 1.0
assi (resi 162 and name HB# ) (resi 161 and name HG2# ) 4.0 4.0 2.5
assi (resi 162 and name HG# ) (resi 162 and name HB# ) 3.4 3.4 1.0
assi (resi 162 and name HG#) (resi 161 and name HG2# ) 3.4 3.4 2.5
assi (resi 162 and name HG# ) (resi 162 and name HB# ) 3.4 3.4 1.0
assi (resi 162 and name HG# ) (resi 161 and name HG2#) 3.4 3.4 2.5
assi (resi 165 and name HB# ) (resi 165 and name HG ) 4.0 4.0 1.0
assi (resi 165 and name HD1# ) (resi 165 and name HB# ) 2.9 2.9 1.0
assi (resi 165 and name HD1# ) (resi 165 and name HG ) 2.4 2.4 1.0
assi (resi 165 and name HD2# ) (resi 165 and name HB# ) 2.4 2.4 1.0
assi (resi 165 and name HD2# ) (resi 165 and name HG ) 2.9 2.9 1.0
assi (resi 167 and name HD2# ) (resi 167 and name HB# ) 2.4 2.4 1.0
assi (resi 167 and name HD2#) (resi 167 and name HD2#) 2.4 2.4 0.0
assi (resi 167 and name HD2# ) (resi 167 and name HB# ) 2.4 2.4 1.0
assi (resi 167 and name HD2# ) (resi 168 and name HB ) 3.4 3.4 0.0
assi (resi 168 and name HB ) (resi 168 and name HG2# ) 3.4 3.4 1.5
assi (resi 168 and name HB ) (resi 168 and name HG1# ) 3.4 3.4 1.5
assi (resi 168 and name HG2#)(resi 168 and name HG1#) 2.4 2.4 2.0
assi (resi 168 and name HG1# ) (resi 169 and name HE# ) 3.4 3.4 3.5
assi (resi 173 and name HG2# ) (resi 173 and name HB ) 2.4 2.4 1.5
assi (resi 173 and name HD# ) (resi 173 and name HB ) 2.4 2.4 0.0
assi (resi 173 and name HD#) (resi 173 and name HG2# ) 2.4 2.4 1.5
assi (resi 174 and name HB# ) (resi 174 and name HB# ) 3.4 3.4 1.0
assi (resi 174 and name HB# ) (resi 174 and name HD## ) 2.9 2.9 1.5
assi (resi 174 and name HD1# ) (resi 174 and name HB# ) 2.9 2.9 1.0
assi (resi 174 and name HD2# ) (resi 174 and name HB# ) 2.9 2.9 1.0
assi (resi 177 and name HB# ) (resi 177 and name HD2# ) 3.4 3.4 1.0
assi (resi 177 and name HB# ) (resi 177 and name HD1# ) 2.9 2.9 1.0
assi (resi 177 and name HG ) (resi 177 and name HD2# ) 2.4 2.4 1.0
assi (resi 177 and name HG ) (resi 177 and name HD1# ) 2.9 2.9 1.0
```

```
assi (resi 177 and name HD2# ) (resi 177 and name HB# ) 2.4 2.4 1.0
assi (resi 179 and name HB# ) (resi 179 and name HD# ) 2.9 2.9 1.0
assi (resi 179 and name HB# ) (resi 179 and name HB# ) 2.4 2.4 1.0

!I+2 TO I+4 CORRELATIONS FROM HN OR HA TO HN OR HA IN
!HELICAL REGIONS

assi (resi  16 and name HN ) (resi  14 and name HN ) 5.5 5.5 0.0
assi (resi  19 and name HN ) (resi  16 and name HA ) 3.4 3.4 0.0
assi (resi  21 and name HN ) (resi  18 and name HA ) 4.0 4.0 0.0
assi (resi  22 and name HN ) (resi  19 and name HA ) 4.0 4.0 0.0
assi (resi  22 and name HN ) (resi  18 and name HA ) 5.5 5.5 0.0
assi (resi  24 and name HN ) (resi  22 and name HA ) 3.4 3.4 0.0
assi (resi  25 and name HN ) (resi  22 and name HA ) 2.9 2.9 0.0
assi (resi  25 and name HN ) (resi  21 and name HA ) 3.4 3.4 0.0
assi (resi  88 and name HN ) (resi  90 and name HN ) 5.5 5.5 0.0
assi (resi  94 and name HN ) (resi  91 and name HA ) 3.4 3.4 0.0
assi (resi  95 and name HN ) (resi  92 and name HA ) 2.9 2.9 0.0
assi (resi  97 and name HN ) (resi  93 and name HA ) 4.0 4.0 0.0
assi (resi 100 and name HN ) (resi  98 and name HN ) 5.5 5.5 0.0
assi (resi 101 and name HN ) (resi  97 and name HA ) 4.0 4.0 0.0
assi (resi 101 and name HN ) (resi  98 and name HA ) 3.4 3.4 0.0
assi (resi 101 and name HD# ) (resi  98 and name HA ) 3.4 3.4 0.0
assi (resi 141 and name HN ) (resi 143 and name HN ) 4.0 4.0 0.0
assi (resi 145 and name HN ) (resi 142 and name HA ) 3.4 3.4 0.0
assi (resi 148 and name HN ) (resi 145 and name HA ) 2.4 2.4 0.0
assi (resi 149 and name HN ) (resi 146 and name HA ) 2.9 2.9 0.0
assi (resi 165 and name HN ) (resi 167 and name HN ) 5.5 5.5 0.0
assi (resi 167 and name HN ) (resi 165 and name HA ) 3.4 3.4 0.0
assi (resi 168 and name HN ) (resi 165 and name HA ) 2.9 2.9 0.0
assi (resi 170 and name HN ) (resi 167 and name HA ) 2.9 2.9 0.0
assi (resi 172 and name HN ) (resi 169 and name HA ) 3.4 3.4 0.0
assi (resi 175 and name HN ) (resi 173 and name HA ) 4.0 4.0 0.0
assi (resi 176 and name HN ) (resi 173 and name HA ) 3.4 3.4 0.0
assi (resi 177 and name HN ) (resi 174 and name HA ) 3.4 3.4 0.0

!I+2 TO I+4 CORRELATIONS FROM HN OR HA TO SIDECHAIN IN
!HELICAL REGIONS

assi (resi  17 and name HA ) (resi  20 and name HD1# ) 3.4 3.4 1.0
assi (resi  18 and name HN ) (resi  20 and name HD1# ) 4.0 4.0 1.0
assi (resi  18 and name HA ) (resi  21 and name HD# ) 3.4 3.4 0.0
assi (resi  19 and name HN ) (resi  17 and name HG2# ) 3.4 3.4 1.5
assi (resi  19 and name HD1# ) (resi  22 and name HN ) 5.5 5.5 1.0
assi (resi  19 and name HD2# ) (resi  16 and name HA ) 4.0 4.0 1.0
assi (resi  22 and name HN ) (resi  19 and name HB# ) 5.5 5.5 1.0
assi (resi  25 and name HN ) (resi  23 and name HD# ) 5.5 5.5 2.0
assi (resi  25 and name HN ) (resi  21 and name HG2# ) 3.4 3.4 1.5
assi (resi  25 and name HG2# ) (resi  21 and name HA ) 3.4 3.4 1.5
assi (resi  27 and name HN ) (resi  25 and name HG2# ) 2.9 2.9 1.5
assi (resi  88 and name HN ) (resi  90 and name HD# ) 4.0 4.0 2.0
assi (resi  90 and name HN ) (resi  93 and name HG2# ) 4.0 4.0 1.5
assi (resi  92 and name HN ) (resi  97 and name HE1 ) 5.5 5.5 0.0
assi (resi  92 and name HD2# ) (resi  89 and name HA ) 5.5 5.5 0.0
assi (resi 100 and name HN ) (resi  97 and name HB# ) 5.5 5.5 1.0
assi (resi 103 and name HN ) (resi 101 and name HD# ) 5.5 5.5 0.0
assi (resi 140 and name HN ) (resi 142 and name HB# ) 5.5 5.5 1.5
assi (resi 142 and name HA ) (resi 145 and name HD2# ) 4.0 4.0 1.0
assi (resi 143 and name HA ) (resi 146 and name HB# ) 3.4 3.4 1.5
assi (resi 146 and name HN ) (resi 143 and name HA ) 3.4 3.4 0.0
assi (resi 146 and name HA ) (resi 151 and name HB# ) 3.4 3.4 1.5
assi (resi 166 and name HN ) (resi 169 and name HD# ) 2.9 2.9 2.0
assi (resi 166 and name HA ) (resi 169 and name HD# ) 2.9 2.9 2.0
assi (resi 167 and name HN ) (resi 169 and name HD# ) 3.4 3.4 2.0
assi (resi 167 and name HD2# ) (resi 164 and name HA1 ) 2.9 2.9 0.0
assi (resi 167 and name HD2# ) (resi 164 and name HA1 ) 3.4 3.4 0.0
assi (resi 168 and name HG2# ) (resi 165 and name HA ) 2.9 2.9 1.5
assi (resi 172 and name HN ) (resi 169 and name HE# ) 5.5 5.5 2.0
assi (resi 172 and name HN ) (resi 168 and name HG2# ) 3.4 3.4 1.5
assi (resi 172 and name HN ) (resi 168 and name HG1# ) 5.5 5.5 1.5
assi (resi 173 and name HA ) (resi 176 and name HB# ) 3.4 3.4 1.5
assi (resi 174 and name HD1# ) (resi 171 and name HA ) 2.9 2.9 1.0
assi (resi 174 and name HD2# ) (resi 171 and name HA ) 3.4 3.4 1.0
assi (resi 175 and name HB# ) (resi 172 and name HA ) 3.4 3.4 1.5
assi (resi 177 and name HD2# ) (resi 174 and name HA ) 3.4 3.4 1.0
assi (resi 177 and name HD1# ) (resi 174 and name HA ) 3.4 3.4 1.0

!I+2 TO I+4 CORRELATIONS FROM SIDECHAIN TO SIDECHAIN IN
! HELICAL REGIONS

assi (resi  17 and name HG2#) (resi  20 and name HD1#) 2.4 2.4 2.5
assi (resi  23 and name HN ) (resi  19 and name HD## ) 3.4 3.4 2.9
assi (resi  25 and name HB ) (resi  21 and name HG2# ) 5.5 5.5 1.5
assi (resi  25 and name HG2# ) (resi  21 and name HG2#) 2.4 2.4 3.0
assi (resi  88 and name HB# ) (resi  91 and name HG# ) 3.4 3.4 2.0
assi (resi  88 and name HB# ) (resi  91 and name HG# ) 2.4 2.4 2.0
assi (resi  88 and name HB# ) (resi  91 and name HG# ) 4.0 4.0 2.0
assi (resi  88 and name HB# ) (resi  91 and name HB# ) 2.9 2.9 2.0
assi (resi  92 and name HD2# ) (resi  97 and name HE1 ) 5.5 5.5 0.0
assi (resi  92 and name HD2# ) (resi  97 and name HE1 ) 5.5 5.5 0.0
assi (resi  97 and name HE1 ) (resi  92 and name HB# ) 3.4 3.4 1.0
assi (resi  97 and name HE1 ) (resi  93 and name HG1# ) 4.0 4.0 1.5
assi (resi  98 and name HG1# ) (resi 101 and name HD# ) 2.9 2.9 1.5
assi (resi  98 and name HG2# ) (resi 101 and name HD# ) 4.0 4.0 1.5
assi (resi 101 and name HD# ) (resi  98 and name HG1# ) 2.4 2.4 1.5
assi (resi 102 and name HG2#) (resi  98 and name HG1#) 2.4 2.4 3.0
assi (resi 103 and name HN ) (resi  98 and name HG1# ) 2.9 2.9 1.5
assi (resi 146 and name HB# ) (resi 151 and name HB# ) 2.4 2.4 3.0
assi (resi 146 and name HB# ) (resi 142 and name HB# ) 2.9 2.9 3.0
assi (resi 149 and name HD1#) (resi 145 and name HD2#) 2.9 2.9 2.0
assi (resi 149 and name HD2#) (resi 145 and name HD2#) 4.0 4.0 2.0
assi (resi 165 and name HD1# ) (resi 169 and name HE# ) 2.9 2.9 3.0
assi (resi 165 and name HD1# ) (resi 169 and name HD# ) 2.9 2.9 3.0
assi (resi 165 and name HD1#)(resi 168 and name HG1#) 4.0 4.0 2.5
assi (resi 165 and name HD2# ) (resi 169 and name HE# ) 2.9 2.9 3.0
assi (resi 165 and name HD2# ) (resi 169 and name HD# ) 2.9 2.9 3.0
assi (resi 165 and name HD2#)(resi 168 and name HG1#) 5.5 5.5 2.5
assi (resi 172 and name HB# ) (resi 169 and name HE# ) 4.0 4.0 3.5
assi (resi 172 and name HB# ) (resi 168 and name HG1# ) 4.0 4.0 3.0

!HN OR HA TO HN OR HA CORRELATIONS BETWEEN BETA
!STRANDS

assi (resi   3 and name HN ) (resi  52 and name HN ) 4.0 4.0 0.0
assi (resi   4 and name HN ) (resi  54 and name HN ) 5.5 5.5 0.0
assi (resi   4 and name HN ) (resi  53 and name HA ) 2.4 2.4 0.0
assi (resi   7 and name HN ) (resi  79 and name HN ) 4.0 4.0 0.0
assi (resi   7 and name HN ) (resi  78 and name HA ) 2.9 2.9 0.0
assi (resi  42 and name HN ) (resi  53 and name HN ) 3.4 3.4 0.0
assi (resi  44 and name HN ) (resi  52 and name HA ) 2.9 2.9 0.0
assi (resi  46 and name HN ) (resi  49 and name HN ) 2.9 2.9 0.0
assi (resi  49 and name HN ) (resi  45 and name HA ) 3.4 3.4 0.0
assi (resi  51 and name HN ) (resi  44 and name HN ) 3.4 3.4 0.0
assi (resi  55 and name HN ) (resi  41 and name HA ) 4.0 4.0 0.0
assi (resi  56 and name HN ) (resi   7 and name HA ) 3.4 3.4 0.0
assi (resi  76 and name HN ) (resi   5 and name HN ) 5.5 5.5 0.0
assi (resi  76 and name HN ) (resi   5 and name HA ) 4.0 4.0 0.0
assi (resi  77 and name HN ) (resi   5 and name HN ) 5.5 5.5 0.0
assi (resi  77 and name HN ) (resi   6 and name HA ) 3.4 3.4 0.0
assi (resi  78 and name HN ) (resi 110 and name HA ) 3.4 3.4 0.0
assi (resi  79 and name HN ) (resi   8 and name HA ) 2.9 2.9 0.0
assi (resi  80 and name HN ) (resi 112 and name HA ) 4.0 4.0 0.0
assi (resi  82 and name HN ) (resi 113 and name HN ) 3.4 3.4 0.0
assi (resi 110 and name HN ) (resi 152 and name HN ) 3.4 3.4 0.0
assi (resi 111 and name HN ) (resi  78 and name HN ) 4.0 4.0 0.0
assi (resi 112 and name HN ) (resi 153 and name HN ) 3.4 3.4 0.0
assi (resi 112 and name HN ) (resi 154 and name HA ) 3.4 3.4 0.0
assi (resi 113 and name HN ) (resi  81 and name HA ) 3.4 3.4 0.0
assi (resi 114 and name HN ) (resi 156 and name HA ) 3.4 3.4 0.0
assi (resi 153 and name HN ) (resi 111 and name HA ) 2.9 2.9 0.0
assi (resi 155 and name HN ) (resi 113 and name HA ) 4.0 4.0 0.0

!HN OR HA TO SIDECHAIN CORRELATIONS BETWEEN BETA
!STRANDS

assi (resi   1 and name HN ) (resi  50 and name HD# ) 3.4 3.4 1.0
assi (resi   3 and name HN ) (resi  52 and name HB ) 3.4 3.4 0.0
assi (resi   4 and name HN ) (resi  52 and name HB ) 3.4 3.4 0.0
assi (resi   4 and name HN ) (resi  52 and name HG2# ) 2.4 2.4 1.5
assi (resi   4 and name HG2# ) (resi  53 and name HA ) 5.5 5.5 1.5
assi (resi   4 and name HD# ) (resi  53 and name HA ) 5.5 5.5 0.0
assi (resi   6 and name HN ) (resi  55 and name HD## ) 2.9 2.9 2.9
assi (resi   7 and name HN ) (resi  77 and name HG2# ) 2.9 2.9 1.5
assi (resi   8 and name HG1# ) (resi  57 and name HA ) 3.4 3.4 1.5
assi (resi   9 and name HN ) (resi  79 and name HB# ) 2.9 2.9 1.0
assi (resi  41 and name HB# ) (resi  54 and name HN ) 4.0 4.0 1.5
assi (resi  41 and name HB# ) (resi  53 and name HA ) 2.9 2.9 1.5
assi (resi  44 and name HN ) (resi  51 and name HD# ) 5.5 5.5 2.0
assi (resi  44 and name HN ) (resi  51 and name HB# ) 4.0 4.0 1.0
assi (resi  45 and name HA ) (resi  50 and name HD# ) 5.5 5.5 1.0
assi (resi  45 and name HE# ) (resi  50 and name HA ) 3.4 3.4 1.5
assi (resi  46 and name HN ) (resi  51 and name HD# ) 3.4 3.4 2.0
assi (resi  46 and name HN ) (resi  51 and name HE# ) 3.4 3.4 2.0
assi (resi  51 and name HN ) (resi  44 and name HG2# ) 2.9 2.9 1.5
assi (resi  52 and name HG2# ) (resi  41 and name HA ) 4.0 4.0 1.5
assi (resi  54 and name HN ) (resi   4 and name HB ) 4.0 4.0 0.0
assi (resi  54 and name HA# ) (resi  41 and name HB# ) 4.0 4.0 2.5
assi (resi  55 and name HN ) (resi   5 and name HD# ) 4.0 4.0 1.0
assi (resi  74 and name HN ) (resi  75 and name HG2# ) 5.5 5.5 1.5
assi (resi  76 and name HN ) (resi   4 and name HG2# ) 4.0 4.0 1.5
assi (resi  77 and name HN ) (resi   6 and name HB# ) 5.5 5.5 1.0
assi (resi  79 and name HN ) (resi   7 and name HB ) 5.5 5.5 0.0
assi (resi  79 and name HN ) (resi   8 and name HG2# ) 2.9 2.9 1.5
assi (resi  80 and name HN ) (resi 110 and name HD# ) 3.4 3.4 2.0
assi (resi  81 and name HN ) (resi   9 and name HG2# ) 2.4 2.4 1.5
assi (resi  82 and name HN ) (resi 112 and name HD1# ) 4.0 4.0 1.0
assi (resi  82 and name HN ) (resi 113 and name HG2# ) 3.4 3.4 1.5
assi (resi  82 and name HA ) (resi 112 and name HD1# ) 4.0 4.0 1.0
```

```
assi (resi 110 and name HN ) (resi 152 and name HB ) 3.4 3.4 0.0
assi (resi 110 and name HN ) (resi 152 and name HG2# ) 2.4 2.4 1.5
assi (resi 110 and name HA ) (resi  78 and name HB# ) 5.5 5.5 1.0
assi (resi 111 and name HN ) (resi  78 and name HB# ) 2.9 2.9 1.0
assi (resi 111 and name HG ) (resi 153 and name HN ) 5.5 5.5 0.0
assi (resi 111 and name HD2# ) (resi 153 and name HN ) 3.4 3.4 1.0
assi (resi 113 and name HG2# ) (resi  81 and name HA ) 4.0 4.0 1.5
assi (resi 113 and name HG1# ) (resi  81 and name HA ) 5.5 5.5 1.5
assi (resi 114 and name HN ) (resi 154 and name HD# ) 4.0 4.0 2.0
assi (resi 115 and name HN ) (resi  84 and name HG2# ) 2.4 2.4 1.5
assi (resi 151 and name HN ) (resi 110 and name HD# ) 4.0 4.0 2.0
assi (resi 151 and name HA ) (resi 110 and name HD# ) 2.9 2.9 2.0
assi (resi 152 and name HN ) (resi 110 and name HB# ) 2.4 2.4 1.0
assi (resi 152 and name HN ) (resi 111 and name HG ) 3.4 3.4 0.0
assi (resi 152 and name HN ) (resi 111 and name HD## ) 2.9 2.9 2.9
assi (resi 152 and name HG1# ) (resi 112 and name HN ) 4.0 4.0 1.5
assi (resi 153 and name HN ) (resi 110 and name HB# ) 3.4 3.4 1.0

!SIDECHAIN TO SIDECHAIN CORRELATIONS BETWEEN BETA
!STRANDS

assi (resi   4 and name HG2# ) (resi  51 and name HB# ) 3.4 3.4 2.5
assi (resi   4 and name HD# ) (resi  51 and name HD# ) 3.4 3.4 2.0
assi (resi   4 and name HD# ) (resi  51 and name HE# ) 4.0 4.0 2.0
assi (resi   4 and name HD# ) (resi  51 and name HB# ) 3.4 3.4 1.0
assi (resi   8 and name HG2# ) (resi  79 and name HD1# ) 2.9 2.9 2.5
assi (resi   8 and name HG1# ) (resi  79 and name HD1# ) 3.4 3.4 2.5
assi (resi   9 and name HG2# ) (resi  80 and name HB ) 5.5 5.5 1.5
assi (resi  41 and name HB# ) (resi  52 and name HB ) 2.4 2.4 1.5
assi (resi  43 and name HB ) (resi  50 and name HB# ) 3.4 3.4 1.0
assi (resi  43 and name HG2# ) (resi  50 and name HB# ) 2.4 2.4 2.5
assi (resi  44 and name HG2# ) (resi  51 and name HB# ) 2.9 2.9 2.5
assi (resi  45 and name HE# ) (resi  51 and name HD# ) 5.5 5.5 3.5
assi (resi  45 and name HE# ) (resi  51 and name HE# ) 5.5 5.5 3.5
assi (resi  46 and name HG2# ) (resi  51 and name HD# ) 3.4 3.4 3.5
assi (resi  46 and name HG2# ) (resi  51 and name HE# ) 2.9 2.9 3.5
assi (resi  46 and name HD# ) (resi  51 and name HE# ) 4.0 4.0 2.0
assi (resi  77 and name HB ) (resi   6 and name HB# ) 4.0 4.0 1.0
assi (resi 111 and name HD1#)(resi 152 and name HG##) 2.9 2.9 3.9
assi (resi 111 and name HD2#)(resi 152 and name HG##) 4.0 4.0 3.9
assi (resi 112 and name HG ) (resi 154 and name HD# ) 3.4 3.4 2.0
assi (resi 112 and name HD1# ) (resi  82 and name HD# ) 2.9 2.9 3.0
assi (resi 112 and name HD1# ) (resi 154 and name HD# ) 2.9 2.9 3.0
assi (resi 112 and name HD1# ) (resi  82 and name HB# ) 4.0 4.0 2.0
assi (resi 112 and name HD1# ) (resi  80 and name HB ) 3.4 3.4 1.0
assi (resi 112 and name HD2# ) (resi  82 and name HD# ) 3.4 3.4 3.0
assi (resi 112 and name HD2# ) (resi 154 and name HD# ) 2.9 2.9 3.0
assi (resi 112 and name HD2# ) (resi 154 and name HB# ) 5.5 5.5 2.0
assi (resi 151 and name HB# ) (resi 110 and name HD# ) 2.4 2.4 3.5
assi (resi 152 and name HB ) (resi 111 and name HD2# ) 3.4 3.4 1.0
assi (resi 152 and name HG2# ) (resi 110 and name HB# ) 3.4 3.4 2.5
assi (resi 152 and name HG2# ) (resi 110 and name HB# ) 2.9 2.9 2.5
assi (resi 152 and name HG2# ) (resi 111 and name HG ) 2.9 2.9 1.5
assi (resi 152 and name HG2#)(resi 111 and name HD1#) 2.9 2.9 2.5

!I TO I+2 CORRELATIONS WITHIN BETA STRAND

assi (resi  45 and name HG# ) (resi  43 and name HG2# ) 2.9 2.9 2.5
assi (resi  45 and name HG# ) (resi  43 and name HG2# ) 3.4 3.4 2.5
assi (resi  45 and name HE# ) (resi  43 and name HG2# ) 2.9 2.9 3.0
assi (resi  46 and name HD# ) (resi  44 and name HG## ) 2.9 2.9 2.9
assi (resi  49 and name HN ) (resi  51 and name HD# ) 4.0 4.0 2.0
assi (resi  49 and name HN ) (resi  51 and name HE# ) 2.9 2.9 2.0
assi (resi  49 and name HB# ) (resi  51 and name HE# ) 3.4 3.4 3.0
assi (resi  49 and name HG# ) (resi  51 and name HE# ) 5.5 5.5 3.0
assi (resi  49 and name HG# ) (resi  51 and name HE# ) 4.0 4.0 3.0
assi (resi  55 and name HN ) (resi  53 and name HA ) 5.5 5.5 0.0
assi (resi  80 and name HG2# ) (resi  78 and name HB# ) 3.4 3.4 2.5
assi (resi 112 and name HN ) (resi 110 and name HD# ) 3.4 3.4 2.0
assi (resi 112 and name HB# ) (resi 110 and name HE# ) 4.0 4.0 3.0
assi (resi 112 and name HB# ) (resi 110 and name HE# ) 5.5 5.5 3.0

!LONG RANGE CONSTRAINTS FROM HN OR HA TO HN OR HA

assi (resi  11 and name HN ) (resi  97 and name HE1 ) 4.0 4.0 0.0
assi (resi  11 and name HN ) (resi  92 and name HD2# ) 5.5 5.5 0.0
assi (resi 112 and name HN ) (resi 152 and name HN ) 3.4 3.4 0.0
assi (resi 159 and name HN ) (resi  19 and name HA ) 3.4 3.4 0.0

!LONG RANGE CONSTRAINTS FROM HN OR HA TO SIDECHAIN

assi (resi   4 and name HG2# ) (resi 173 and name HA ) 5.5 5.5 1.5
assi (resi   4 and name HD# ) (resi 173 and name HA ) 4.0 4.0 0.0
assi (resi   9 and name HG1# ) (resi  97 and name HA ) 3.4 3.4 1.5
assi (resi   9 and name HG2# ) (resi  97 and name HA ) 2.9 2.9 1.5
assi (resi  15 and name HN ) (resi  85 and name HG2# ) 4.0 4.0 1.5
assi (resi  23 and name HN ) (resi 169 and name HE# ) 3.4 3.4 2.0
assi (resi  26 and name HA ) (resi 159 and name HB# ) 4.0 4.0 1.5
assi (resi  53 and name HD2# ) (resi 169 and name HA ) 5.5 5.5 1.0
assi (resi  85 and name HG2# ) (resi 120 and name HN ) 5.5 5.5 1.5
assi (resi  89 and name HA ) (resi 129 and name HD## ) 2.9 2.9 2.9
assi (resi  90 and name HN ) (resi  82 and name HD# ) 3.4 3.4 2.0
assi (resi  90 and name HN ) (resi 137 and name HD# ) 3.4 3.4 0.0
assi (resi  90 and name HA ) (resi  82 and name HD# ) 5.5 5.5 2.0
assi (resi  93 and name HN ) (resi 145 and name HD2# ) 4.0 4.0 1.0
assi (resi  93 and name HA ) (resi  80 and name HG2# ) 5.5 5.5 1.5
assi (resi  93 and name HA ) (resi  80 and name HG1# ) 3.4 3.4 1.5
assi (resi  94 and name HN ) (resi 145 and name HD2# ) 4.0 4.0 1.0
assi (resi 101 and name HD# ) (resi  80 and name HN ) 5.5 5.5 0.0
assi (resi 109 and name HA ) (resi 152 and name HG2# ) 3.4 3.4 1.5
assi (resi 111 and name HD1# ) (resi 172 and name HN ) 3.4 3.4 1.0
assi (resi 111 and name HD1# ) (resi 171 and name HA ) 3.4 3.4 1.0
assi (resi 111 and name HD1# ) (resi 172 and name HA ) 2.9 2.9 1.0
assi (resi 111 and name HD1# ) (resi 168 and name HA ) 4.0 4.0 1.0
assi (resi 111 and name HD2# ) (resi 171 and name HA ) 3.4 3.4 1.0
assi (resi 111 and name HD2# ) (resi 172 and name HA ) 3.4 3.4 1.0
assi (resi 111 and name HD2# ) (resi 168 and name HA ) 3.4 3.4 1.0
assi (resi 112 and name HD2# ) (resi 146 and name HN ) 4.0 4.0 1.0
assi (resi 112 and name HD2# ) (resi 146 and name HA ) 3.4 3.4 1.0
assi (resi 116 and name HE2# ) (resi  18 and name HN ) 3.4 3.4 0.0
assi (resi 116 and name HE2# ) (resi  18 and name HA ) 5.5 5.5 0.0
assi (resi 116 and name HE2# ) (resi  18 and name HN ) 3.4 3.4 0.0
assi (resi 116 and name HE2# ) (resi  17 and name HN ) 4.0 4.0 0.0
assi (resi 118 and name HN ) (resi 160 and name HD2# ) 3.4 3.4 1.0
assi (resi 125 and name HA ) (resi  85 and name HG2# ) 4.0 4.0 1.5
assi (resi 137 and name HD# ) (resi  84 and name HA ) 2.9 2.9 0.0
assi (resi 137 and name HG2# ) (resi  84 and name HA ) 5.5 5.5 1.5
assi (resi 141 and name HN ) (resi  90 and name HE# ) 4.0 4.0 2.0
assi (resi 142 and name HN ) (resi  90 and name HE# ) 3.4 3.4 2.0
assi (resi 142 and name HA ) (resi  90 and name HE# ) 2.9 2.9 2.0
assi (resi 143 and name HN ) (resi 154 and name HE# ) 4.0 4.0 2.0
assi (resi 145 and name HN ) (resi  90 and name HE# ) 2.9 2.9 2.0
assi (resi 145 and name HD1# ) (resi  91 and name HA ) 4.0 4.0 1.0
assi (resi 146 and name HN ) (resi 154 and name HE# ) 4.0 4.0 2.0
assi (resi 146 and name HN ) (resi  90 and name HE# ) 4.0 4.0 2.0
assi (resi 146 and name HB# ) (resi 154 and name HA ) 4.0 4.0 1.5
assi (resi 148 and name HN ) (resi 145 and name HD1# ) 3.4 3.4 1.0
assi (resi 149 and name HD1# ) (resi  94 and name HA ) 3.4 3.4 1.0
assi (resi 149 and name HD2# ) (resi  94 and name HA ) 2.9 2.9 1.0
assi (resi 151 and name HB# ) (resi 112 and name HA ) 5.5 5.5 1.5
assi (resi 152 and name HG2# ) (resi 175 and name HA ) 4.0 4.0 1.5
assi (resi 152 and name HG1# ) (resi 175 and name HA ) 2.9 2.9 1.5
assi (resi 155 and name HG2# ) (resi 168 and name HA ) 2.9 2.9 1.5
assi (resi 157 and name HA ) (resi 164 and name HA# ) 5.5 5.5 1.0
assi (resi 159 and name HA ) (resi  19 and name HD2# ) 3.4 3.4 1.0
assi (resi 159 and name HB# ) (resi  19 and name HA ) 2.9 2.9 1.5
assi (resi 160 and name HN ) (resi  28 and name HD# ) 2.9 2.9 2.0
assi (resi 160 and name HA ) (resi  28 and name HD# ) 5.5 5.5 2.0
assi (resi 161 and name HN ) (resi 118 and name HB# ) 5.5 5.5 1.0
assi (resi 165 and name HN ) (resi 157 and name HB# ) 4.0 4.0 1.0
assi (resi 165 and name HD1# ) (resi  23 and name HN ) 5.5 5.5 1.0
assi (resi 165 and name HD2# ) (resi  23 and name HN ) 4.0 4.0 1.0
assi (resi 168 and name HN ) (resi 113 and name HG2# ) 5.5 5.5 1.5
assi (resi 171 and name HN ) (resi 111 and name HD## ) 2.9 2.9 2.9
assi (resi 175 and name HN ) (resi 152 and name HG1# ) 4.0 4.0 1.5

!LONG RANGE CONSTRAINTS FROM SIDECHAIN TO SIDECHAIN

assi (resi   2 and name HE2# ) (resi 177 and name HD2# ) 3.4 3.4 1.0
assi (resi   2 and name HE2# ) (resi 177 and name HD2# ) 2.9 2.9 1.0
assi (resi   4 and name HG2# ) (resi 176 and name HB# ) 2.4 2.4 3.0
assi (resi   4 and name HD# ) (resi 173 and name HG2# ) 2.9 2.9 1.5
assi (resi   9 and name HG2# ) (resi  97 and name HB# ) 3.4 3.4 2.5
assi (resi  12 and name HN ) (resi  97 and name HZ2 ) 3.4 3.4 0.0
assi (resi  17 and name HG2# ) (resi  32 and name HB# ) 2.9 2.9 2.5
assi (resi  19 and name HG ) (resi 169 and name HE# ) 3.4 3.4 2.0
assi (resi  19 and name HD1# ) (resi 169 and name HE# ) 2.4 2.4 3.0
assi (resi  19 and name HD1# ) (resi  79 and name HD1# ) 4.0 4.0 2.0
assi (resi  19 and name HD1#) (resi 168 and name HG1#) 2.9 2.9 2.5
assi (resi  19 and name HD2# ) (resi 169 and name HE# ) 2.9 2.9 3.0
assi (resi  19 and name HD2# ) (resi  79 and name HD1# ) 3.4 3.4 2.0
assi (resi  19 and name HD2#) (resi 168 and name HG1#) 3.4 3.4 2.5
assi (resi  21 and name HG2# ) (resi  32 and name HB# ) 3.4 3.4 2.5
assi (resi  21 and name HD# ) (resi  17 and name HG2# ) 2.9 2.9 1.5
assi (resi  22 and name HB# ) (resi 165 and name HD2# ) 2.9 2.9 2.0
assi (resi  23 and name HA ) (resi 165 and name HD2# ) 2.9 2.9 1.0
assi (resi  42 and name HG1# ) (resi  23 and name HD# ) 2.9 2.9 3.5
assi (resi  42 and name HG1# ) (resi  24 and name HB ) 2.9 2.9 1.5
assi (resi  42 and name HG2# ) (resi  23 and name HD# ) 2.9 2.9 3.5
assi (resi  42 and name HG2# ) (resi  24 and name HB ) 2.9 2.9 1.5
assi (resi  44 and name HG1# ) (resi 169 and name HB# ) 4.0 4.0 2.5
assi (resi  46 and name HD# ) (resi 170 and name HB# ) 3.4 3.4 1.0
assi (resi  53 and name HD2# ) (resi 169 and name HE# ) 2.9 2.9 3.0
assi (resi  53 and name HD2# ) (resi 169 and name HD# ) 2.9 2.9 3.0
assi (resi  53 and name HD2# ) (resi 169 and name HB# ) 5.5 5.5 2.0
assi (resi  55 and name HD2# ) (resi  23 and name HB# ) 4.0 4.0 2.0
assi (resi  55 and name HD2# ) (resi  23 and name HB# ) 3.4 3.4 2.0
assi (resi  77 and name HG1# ) (resi 172 and name HB# ) 2.4 2.4 3.0
assi (resi  79 and name HD1# ) (resi 169 and name HE# ) 2.9 2.9 3.0
assi (resi  79 and name HD1#) (resi 168 and name HG1#) 2.9 2.9 2.5
assi (resi  79 and name HD2# ) (resi 169 and name HE# ) 5.5 5.5 3.0
assi (resi  79 and name HD2# ) (resi 172 and name HB# ) 2.4 2.4 2.5
assi (resi  79 and name HD2#) (resi 168 and name HG1#) 2.9 2.9 2.5
assi (resi  80 and name HG2#) (resi 101 and name HG2#) 3.4 3.4 3.0
assi (resi  85 and name HG1#) (resi 125 and name HG2#) 2.9 2.9 3.0
assi (resi  94 and name HG# ) (resi 145 and name HD2# ) 4.0 4.0 2.0
assi (resi  94 and name HD# ) (resi 145 and name HD2# ) 3.4 3.4 2.0
assi (resi  94 and name HE# ) (resi 145 and name HD2# ) 2.9 2.9 2.0
assi (resi  98 and name HG1# ) (resi 110 and name HE# ) 2.9 2.9 3.5
assi (resi  98 and name HG2# ) (resi 110 and name HE# ) 2.9 2.9 3.5
assi (resi 101 and name HG2#) (resi   9 and name HG1# ) 2.9 2.9 3.0
assi (resi 101 and name HG2# ) (resi   9 and name HG2#) 2.4 2.4 3.0
assi (resi 101 and name HD# ) (resi 110 and name HD# ) 2.9 2.9 2.0
assi (resi 102 and name HG2# ) (resi 110 and name HE# ) 2.9 2.9 3.5
assi (resi 102 and name HG2# ) (resi 110 and name HB# ) 5.5 5.5 2.5
assi (resi 111 and name HD1# ) (resi 172 and name HB# ) 2.4 2.4 2.5
assi (resi 111 and name HD1# ) (resi 175 and name HB# ) 2.4 2.4 2.5
assi (resi 111 and name HD1#)(resi 168 and name HG1#) 3.4 3.4 2.5
assi (resi 111 and name HD2# ) (resi 175 and name HB# ) 2.4 2.4 2.5
assi (resi 111 and name HD2#)(resi 168 and name HG1#) 4.0 4.0 2.5
assi (resi 112 and name HD1#)(resi 145 and name HD1# ) 2.4 2.4 2.0
assi (resi 112 and name HD1#) (resi 145 and name HD2#) 2.9 2.9 2.0
assi (resi 112 and name HD2# ) (resi 146 and name HB# ) 2.9 2.9 2.5
assi (resi 112 and name HD2#) (resi 149 and name HD1#) 2.4 2.4 2.0
assi (resi 112 and name HD2#) (resi 145 and name HD2#) 2.9 2.9 2.0
assi (resi 113 and name HG2#)(resi 168 and name HG1#) 2.9 2.9 3.0
assi (resi 113 and name HG1# ) (resi 168 and name HB ) 3.4 3.4 1.5
assi (resi 113 and name HG1#)(resi 168 and name HG1#) 2.9 2.9 3.0
assi (resi 125 and name HG2#)(resi  85 and name HG2# ) 2.4 2.4 3.0
assi (resi 137 and name HN ) (resi  84 and name HB ) 5.5 5.5 0.0
assi (resi 137 and name HN ) (resi  84 and name HG2# ) 2.9 2.9 1.5
assi (resi 137 and name HD# ) (resi  90 and name HE# ) 3.4 3.4 2.0
assi (resi 137 and name HD# ) (resi  90 and name HD# ) 3.4 3.4 2.0
assi (resi 137 and name HD# ) (resi  90 and name HB# ) 3.4 3.4 1.0
assi (resi 137 and name HD# ) (resi  90 and name HB# ) 3.4 3.4 1.0
assi (resi 137 and name HD#) (resi 141 and name HG2# ) 2.4 2.4 1.5
assi (resi 137 and name HG2# ) (resi  90 and name HE# ) 2.9 2.9 3.5
assi (resi 137 and name HG2# ) (resi  90 and name HB# ) 4.0 4.0 2.5
assi (resi 141 and name HG2# ) (resi  90 and name HE# ) 2.4 2.4 3.5
assi (resi 141 and name HG2# ) (resi  90 and name HD# ) 4.0 4.0 3.5
assi (resi 142 and name HB# ) (resi 154 and name HE# ) 2.9 2.9 3.5
assi (resi 143 and name HA ) (resi 154 and name HB# ) 4.0 4.0 1.0
assi (resi 143 and name HG# ) (resi 154 and name HD# ) 3.4 3.4 3.0
assi (resi 145 and name HD1# ) (resi  90 and name HE# ) 2.9 2.9 3.0
assi (resi 145 and name HD1# ) (resi  94 and name HE# ) 2.9 2.9 2.0
assi (resi 145 and name HD2# ) (resi  90 and name HE# ) 2.9 2.9 3.0
assi (resi 145 and name HD2# ) (resi  90 and name HD# ) 5.5 5.5 3.0
assi (resi 145 and name HD2# ) (resi  94 and name HE# ) 4.0 4.0 2.0
assi (resi 145 and name HD2# ) (resi  93 and name HB ) 4.0 4.0 1.0
assi (resi 146 and name HB# ) (resi 154 and name HB# ) 2.9 2.9 2.5
assi (resi 146 and name HB# ) (resi 112 and name HG ) 2.9 2.9 1.5
assi (resi 149 and name HD1# ) (resi  94 and name HE# ) 5.5 5.5 2.0
assi (resi 149 and name HD2# ) (resi  94 and name HE# ) 2.4 2.4 2.0
assi (resi 152 and name HG1# ) (resi 175 and name HB# ) 2.9 2.9 3.0
assi (resi 155 and name HG1#)(resi 168 and name HG1#) 4.0 4.0 3.0
assi (resi 155 and name HG2#)(resi 168 and name HG1#) 2.9 2.9 3.0
assi (resi 159 and name HB# ) (resi  28 and name HD# ) 2.9 2.9 3.5
assi (resi 159 and name HB# ) (resi  28 and name HE# ) 2.9 2.9 3.5
assi (resi 159 and name HB# ) (resi  22 and name HB# ) 2.9 2.9 2.5
assi (resi 159 and name HB#)(resi  19 and name HD##) 2.9 2.9 4.400
assi (resi 160 and name HD2# ) (resi  28 and name HE# ) 3.4 3.4 3.0
assi (resi 160 and name HD2# ) (resi 118 and name HB# ) 5.5 5.5 2.0
assi (resi 160 and name HD1# ) (resi  28 and name HD# ) 2.9 2.9 3.0
assi (resi 160 and name HD1# ) (resi  28 and name HZ ) 2.9 2.9 1.0
assi (resi 160 and name HD1# ) (resi  28 and name HB# ) 3.4 3.4 2.0
assi (resi 160 and name HD1#)(resi 160 and name HD2# ) 2.4 2.4 1.0
assi (resi 165 and name HD1# ) (resi  23 and name HB# ) 5.5 5.5 2.0
assi (resi 165 and name HD2# ) (resi  23 and name HB# ) 4.0 4.0 2.0
assi (resi 167 and name HD2# ) (resi 155 and name HB ) 2.9 2.9 0.0
assi (resi 167 and name HD2#)(resi 155 and name HG1#) 2.4 2.4 1.5
assi (resi 167 and name HD2# ) (resi 155 and name HB ) 2.9 2.9 0.0
assi (resi 167 and name HD2#)(resi 155 and name HG##) 2.4 2.4 2.9
assi (resi 168 and name HG2#)(resi 113 and name HG1#) 5.5 5.5 3.0
assi (resi 172 and name HB# ) (resi 111 and name HD1# ) 2.4 2.4 2.5
assi (resi 173 and name HG2# ) (resi  51 and name HD# ) 5.5 5.5 3.5
assi (resi 173 and name HG2# ) (resi  51 and name HB# ) 5.5 5.5 2.5
assi (resi 173 and name HD# ) (resi  51 and name HB# ) 3.4 3.4 1.0
assi (resi 176 and name HB# ) (resi   4 and name HD# ) 2.9 2.9 1.5
assi (resi 176 and name HB# ) (resi  77 and name HG1# ) 2.4 2.4 3.0
assi (resi 176 and name HB# ) (resi  77 and name HG2# ) 5.5 5.5 3.0
assi (resi 177 and name HD2# ) (resi  51 and name HE# ) 2.9 2.9 3.0
assi (resi 177 and name HD1# ) (resi  51 and name HE# ) 3.4 3.4 3.0

!MISCELLANEOUS DISTANCES

assi (resi  -2 and name HN ) (resi  51 and name HE# ) 4.0 4.0 2.0
assi (resi   2 and name HN ) (resi  -1 and name HB# ) 3.4 3.4 1.5
assi (resi  18 and name HN ) (resi  28 and name HE# ) 5.5 5.5 2.0
assi (resi  20 and name HD1# ) (resi  23 and name HD# ) 5.5 5.5 3.0
assi (resi  24 and name HG2# ) (resi  40 and name HD# ) 2.9 2.9 3.5
assi (resi  24 and name HG2# ) (resi  23 and name HD# ) 5.5 5.5 3.5
assi (resi  25 and name HG2# ) (resi  40 and name HD# ) 3.4 3.4 3.5
assi (resi  26 and name HN ) (resi  23 and name HD# ) 4.0 4.0 2.0
assi (resi  26 and name HN ) (resi  22 and name HA ) 2.9 2.9 0.0
assi (resi  27 and name HN ) (resi  22 and name HA ) 2.4 2.4 0.0
assi (resi  28 and name HN ) (resi 160 and name HD1# ) 3.4 3.4 1.0
assi (resi  29 and name HA ) (resi  33 and name HG1# ) 2.9 2.9 1.5
assi (resi  30 and name HN ) (resi  28 and name HD# ) 3.4 3.4 2.0
assi (resi  30 and name HN ) (resi  28 and name HB# ) 3.4 3.4 1.0
assi (resi  30 and name HN ) (resi 160 and name HD1# ) 4.0 4.0 1.0
assi (resi  30 and name HA ) (resi  28 and name HZ ) 4.0 4.0 0.0
assi (resi  30 and name HA ) (resi  28 and name HE# ) 3.4 3.4 2.0
assi (resi  30 and name HB# ) (resi  28 and name HZ ) 2.9 2.9 1.0
assi (resi  30 and name HB# ) (resi  28 and name HD# ) 3.4 3.4 3.0
assi (resi  30 and name HB# ) (resi  28 and name HE# ) 3.4 3.4 3.0
assi (resi  35 and name HN ) (resi  33 and name HG2# ) 3.4 3.4 1.5
assi (resi  47 and name HN ) (resi  49 and name HN ) 5.5 5.5 0.0
assi (resi  48 and name HN ) (resi  45 and name HG# ) 4.0 4.0 1.0
assi (resi  48 and name HN ) (resi  45 and name HE# ) 2.9 2.9 1.5
assi (resi  48 and name HN ) (resi  46 and name HG2# ) 4.0 4.0 1.5
assi (resi  48 and name HA# ) (resi  45 and name HE# ) 5.5 5.5 2.5
assi (resi  48 and name HA# ) (resi  45 and name HE# ) 4.0 4.0 2.5
assi (resi  61 and name HE2# ) (resi  63 and name HB# ) 4.0 4.0 1.0
assi (resi  62 and name HN ) (resi  64 and name HN ) 5.5 5.5 0.0
assi (resi  63 and name HN ) (resi  61 and name HG# ) 3.4 3.4 1.0
assi (resi  84 and name HN ) (resi 115 and name HN ) 3.4 3.4 0.0
assi (resi  84 and name HG2# ) (resi 117 and name HD# ) 2.9 2.9 1.5
assi (resi  84 and name HG1#)(resi 137 and name HG2# ) 4.0 4.0 3.0
assi (resi  85 and name HN ) (resi  83 and name HA ) 2.9 2.9 0.0
assi (resi  85 and name HN ) (resi  83 and name HB# ) 4.0 4.0 1.0
assi (resi  86 and name HN ) (resi  84 and name HN ) 4.0 4.0 0.0
assi (resi  86 and name HN ) (resi  90 and name HD# ) 5.5 5.5 2.0
assi (resi  88 and name HN ) (resi  86 and name HA ) 5.5 5.5 0.0
assi (resi  90 and name HN ) (resi  13 and name HN ) 5.5 5.5 0.0
assi (resi 105 and name HN ) (resi 101 and name HG2# ) 4.0 4.0 1.5
assi (resi 107 and name HN ) (resi 105 and name HA ) 4.0 4.0 0.0
assi (resi 115 and name HN ) (resi  83 and name HA ) 2.9 2.9 0.0
assi (resi 115 and name HN ) (resi 113 and name HB ) 5.5 5.5 0.0
assi (resi 116 and name HN ) (resi 158 and name HA ) 4.0 4.0 0.0
assi (resi 116 and name HN ) (resi 159 and name HB# ) 5.5 5.5 1.5
assi (resi 116 and name HN ) (resi  85 and name HG2# ) 3.4 3.4 1.5
assi (resi 116 and name HE2# ) (resi 159 and name HB# ) 4.0 4.0 1.5
assi (resi 117 and name HN ) (resi 158 and name HA ) 3.4 3.4 0.0
assi (resi 117 and name HG2# ) (resi 157 and name HN ) 3.4 3.4 1.5
assi (resi 117 and name HG2# ) (resi 157 and name HA ) 4.0 4.0 1.5
assi (resi 117 and name HD# ) (resi 157 and name HN ) 3.4 3.4 0.0
assi (resi 117 and name HD# ) (resi 157 and name HA ) 3.4 3.4 0.0
assi (resi 118 and name HN ) (resi 158 and name HA ) 3.4 3.4 0.0
assi (resi 118 and name HN ) (resi 120 and name HB# ) 4.0 4.0 1.0
assi (resi 119 and name HD1# ) (resi 122 and name HB# ) 3.4 3.4 2.0
assi (resi 119 and name HD2# ) (resi 122 and name HB# ) 3.4 3.4 2.0
assi (resi 120 and name HN ) (resi 118 and name HN ) 4.0 4.0 0.0
assi (resi 120 and name HA ) (resi  85 and name HG1# ) 2.9 2.9 1.5
assi (resi 121 and name HN ) (resi 119 and name HN ) 4.0 4.0 0.0
assi (resi 121 and name HN ) (resi 119 and name HA ) 3.4 3.4 0.0
assi (resi 122 and name HN ) (resi 124 and name HN ) 4.0 4.0 0.0
assi (resi 122 and name HN ) (resi 119 and name HN ) 4.0 4.0 0.0
assi (resi 122 and name HN ) (resi 119 and name HA ) 2.9 2.9 0.0
assi (resi 122 and name HN ) (resi 119 and name HB# ) 3.4 3.4 1.0
assi (resi 122 and name HN ) (resi 125 and name HG2# ) 4.0 4.0 1.5
assi (resi 124 and name HN ) (resi 122 and name HA ) 4.0 4.0 0.0
assi (resi 124 and name HB# ) (resi 119 and name HA ) 4.0 4.0 1.0
assi (resi 124 and name HB# ) (resi 119 and name HA ) 3.4 3.4 1.0
assi (resi 125 and name HA ) (resi 128 and name HB# ) 2.9 2.9 1.0
assi (resi 125 and name HA ) (resi 128 and name HD# ) 4.0 4.0 1.0
assi (resi 125 and name HG2# ) (resi 128 and name HB# ) 2.9 2.9 2.5
assi (resi 126 and name HB ) (resi 123 and name HA ) 4.0 4.0 0.0
assi (resi 126 and name HG2# ) (resi 123 and name HA ) 2.9 2.9 1.5
assi (resi 126 and name HD# ) (resi 123 and name HA ) 3.4 3.4 0.0
assi (resi 126 and name HD# ) (resi 134 and name HG# ) 3.4 3.4 1.0
assi (resi 126 and name HD# ) (resi 134 and name HB# ) 2.9 2.9 1.0
assi (resi 128 and name HN ) (resi 126 and name HA ) 5.5 5.5 0.0
assi (resi 129 and name HN ) (resi 126 and name HA ) 3.4 3.4 0.0
assi (resi 130 and name HB# ) (resi 127 and name HA ) 2.4 2.4 1.5
assi (resi 130 and name HB# ) (resi 126 and name HA ) 5.5 5.5 1.5
assi (resi 130 and name HB# ) (resi 127 and name HB# ) 3.4 3.4 2.5
assi (resi 130 and name HB# ) (resi 126 and name HG2# ) 2.9 2.9 3.0
assi (resi 132 and name HD2#)(resi 129 and name HD## ) 4.0 4.0 2.9
assi (resi 132 and name HD2#)(resi 129 and name HD## ) 3.4 3.4 2.9
assi (resi 132 and name HB# ) (resi 129 and name HD2# ) 5.5 5.5 2.0
assi (resi 132 and name HB# ) (resi 134 and name HB# ) 3.4 3.4 2.0
assi (resi 132 and name HB# ) (resi 129 and name HD2# ) 5.5 5.5 2.0
assi (resi 133 and name HN ) (resi 130 and name HA ) 2.9 2.9 0.0
assi (resi 133 and name HN ) (resi 130 and name HB# ) 2.9 2.9 1.5
assi (resi 134 and name HN ) (resi 132 and name HN ) 2.9 2.9 0.0
assi (resi 134 and name HN ) (resi 130 and name HA ) 2.9 2.9 0.0
```

```
assi (resi 134 and name HE2# ) (resi 126 and name HD# ) 3.4 3.4 0.0
assi (resi 134 and name HE2# ) (resi 126 and name HD# ) 3.4 3.4 0.0
assi (resi 135 and name HN ) (resi 126 and name HD# ) 4.0 4.0 0.0
assi (resi 137 and name HA ) (resi 141 and name HG2# ) 2.9 2.9 1.5
assi (resi 137 and name HG1# ) (resi 142 and name HB# ) 2.4 2.4 2.5
assi (resi 137 and name HG1# ) (resi 142 and name HB# ) 2.9 2.9 2.5
assi (resi 137 and name HD# ) (resi  84 and name HG2# ) 4.0 4.0 1.5
assi (resi 137 and name HG2#)(resi 141 and name HG2#) 2.4 2.4 3.0
assi (resi 137 and name HG2# ) (resi 142 and name HB# ) 2.9 2.9 3.0
assi (resi 138 and name HN ) (resi 141 and name HN ) 3.4 3.4 0.0
assi (resi 139 and name HA ) (resi 142 and name HB# ) 3.4 3.4 1.5
assi (resi 140 and name HN ) (resi 138 and name HB ) 2.4 2.4 0.0
assi (resi 140 and name HN ) (resi 138 and name HG2# ) 3.4 3.4 1.5
assi (resi 140 and name HN ) (resi 137 and name HG2# ) 5.5 5.5 1.5
assi (resi 141 and name HN ) (resi 137 and name HG2# ) 3.4 3.4 1.5
assi (resi 141 and name HA ) (resi 137 and name HG2# ) 5.5 5.5 1.5
assi (resi 142 and name HN ) (resi 137 and name HG2# ) 2.9 2.9 1.5
assi (resi 150 and name HN ) (resi 146 and name HA ) 2.9 2.9 0.0
assi (resi 157 and name HN ) (resi 117 and name HB ) 3.4 3.4 0.0
assi (resi 157 and name HN ) (resi 117 and name HD# ) 2.9 2.9 0.0
assi (resi 157 and name HA ) (resi 117 and name HD# ) 4.0 4.0 0.0
assi (resi 158 and name HN ) (resi 163 and name HN ) 5.5 5.5 0.0
assi (resi 158 and name HN ) (resi 162 and name HN ) 5.5 5.5 0.0
assi (resi 158 and name HN ) (resi 163 and name HB# ) 5.5 5.5 1.0
assi (resi 158 and name HN ) (resi 117 and name HG2# ) 3.4 3.4 1.5
assi (resi 158 and name HN ) (resi 117 and name HD# ) 3.4 3.4 0.0
assi (resi 159 and name HN ) (resi 165 and name HD1# ) 3.4 3.4 1.0
assi (resi 159 and name HN ) (resi 117 and name HG2# ) 5.5 5.5 1.5
assi (resi 160 and name HN ) (resi 158 and name HA ) 4.0 4.0 0.0
assi (resi 161 and name HN ) (resi 159 and name HN ) 5.5 5.5 0.0
assi (resi 161 and name HN ) (resi 157 and name HA ) 5.5 5.5 0.0
assi (resi 162 and name HN ) (resi 157 and name HA ) 4.0 4.0 0.0
assi (resi 162 and name HN ) (resi 159 and name HA ) 5.5 5.5 0.0
assi (resi 162 and name HN ) (resi 159 and name HB# ) 5.5 5.5 1.5
assi (resi 162 and name HN ) (resi 165 and name HD1# ) 2.9 2.9 1.0
assi (resi 163 and name HN ) (resi 157 and name HA ) 2.9 2.9 0.0
assi (resi 163 and name HN ) (resi 161 and name HB ) 4.0 4.0 0.0
assi (resi 163 and name HN ) (resi 161 and name HG2# ) 2.9 2.9 1.5
assi (resi 164 and name HN ) (resi 157 and name HA ) 4.0 4.0 0.0
assi (resi 165 and name HD1# ) (resi 159 and name HA ) 2.9 2.9 1.0
assi (resi 165 and name HD2# ) (resi 159 and name HA ) 2.9 2.9 1.0
assi (resi 179 and name HB# ) (resi 177 and name HD2# ) 2.9 2.9 2.0
assi (resi 187 and name HN ) (resi 185 and name HB# ) 4.0 4.0 1.0

!HBONDS IN HELICAL REGIONS

assi (resi  20 and name N ) (resi  16 and name O ) 3.3 0.8 0.0
assi (resi  20 and name HN ) (resi  16 and name O ) 2.3 0.5 0.0
assi (resi  21 and name N ) (resi  17 and name O ) 3.3 0.8 0.0
assi (resi  21 and name HN ) (resi  17 and name O ) 2.3 0.5 0.0
assi (resi  22 and name N ) (resi  18 and name O ) 3.3 0.8 0.0
assi (resi  22 and name HN ) (resi  18 and name O ) 2.3 0.5 0.0
assi (resi  23 and name N ) (resi  19 and name O ) 3.3 0.8 0.0
assi (resi  23 and name HN ) (resi  19 and name O ) 2.3 0.5 0.0
assi (resi  24 and name N ) (resi  20 and name O ) 3.3 0.8 0.0
assi (resi  24 and name HN ) (resi  20 and name O ) 2.3 0.5 0.0
assi (resi  95 and name N ) (resi  91 and name O ) 3.3 0.8 0.0
assi (resi  95 and name HN ) (resi  91 and name O ) 2.3 0.5 0.0
assi (resi  96 and name N ) (resi  92 and name O ) 3.3 0.8 0.0
assi (resi  96 and name HN ) (resi  92 and name O ) 2.3 0.5 0.0
assi (resi  97 and name N ) (resi  93 and name O ) 3.3 0.8 0.0
assi (resi  97 and name HN ) (resi  93 and name O ) 2.3 0.5 0.0
assi (resi  98 and name N ) (resi  94 and name O ) 3.3 0.8 0.0
assi (resi  98 and name HN ) (resi  94 and name O ) 2.3 0.5 0.0
assi (resi 100 and name N ) (resi  96 and name O ) 3.3 0.8 0.0
assi (resi 100 and name HN ) (resi  96 and name O ) 2.3 0.5 0.0
assi (resi 101 and name N ) (resi  97 and name O ) 3.3 0.8 0.0
assi (resi 101 and name HN ) (resi  97 and name O ) 2.3 0.5 0.0
assi (resi 102 and name N ) (resi  98 and name O ) 3.3 0.8 0.0
assi (resi 102 and name HN ) (resi  98 and name O ) 2.3 0.5 0.0
assi (resi 145 and name N ) (resi 141 and name O ) 3.3 0.8 0.0
assi (resi 145 and name HN ) (resi 141 and name O ) 2.3 0.5 0.0
assi (resi 146 and name N ) (resi 142 and name O ) 3.3 0.8 0.0
assi (resi 146 and name HN ) (resi 142 and name O ) 2.3 0.5 0.0
assi (resi 147 and name N ) (resi 143 and name O ) 3.3 0.8 0.0
assi (resi 147 and name HN ) (resi 143 and name O ) 2.3 0.5 0.0
assi (resi 169 and name N ) (resi 165 and name O ) 3.3 0.8 0.0
assi (resi 169 and name HN ) (resi 165 and name O ) 2.3 0.5 0.0
assi (resi 170 and name N ) (resi 166 and name O ) 3.3 0.8 0.0
assi (resi 170 and name HN ) (resi 166 and name O ) 2.3 0.5 0.0
assi (resi 171 and name N ) (resi 167 and name O ) 3.3 0.8 0.0
assi (resi 171 and name HN ) (resi 167 and name O ) 2.3 0.5 0.0
assi (resi 172 and name N ) (resi 168 and name O ) 3.3 0.8 0.0
assi (resi 172 and name HN ) (resi 168 and name O ) 2.3 0.5 0.0
assi (resi 173 and name N ) (resi 169 and name O ) 3.3 0.8 0.0
assi (resi 173 and name HN ) (resi 169 and name O ) 2.3 0.5 0.0
assi (resi 174 and name N ) (resi 170 and name O ) 3.3 0.8 0.0
assi (resi 174 and name HN ) (resi 170 and name O ) 2.3 0.5 0.0
assi (resi 175 and name N ) (resi 171 and name O ) 3.3 0.8 0.0
assi (resi 175 and name HN ) (resi 171 and name O ) 2.3 0.5 0.0
assi (resi 176 and name N ) (resi 172 and name O ) 3.3 0.8 0.0
assi (resi 176 and name HN ) (resi 172 and name O ) 2.3 0.5 0.0
assi (resi 177 and name N ) (resi 173 and name O ) 3.3 0.8 0.0
assi (resi 177 and name HN ) (resi 173 and name O ) 2.3 0.5 0.0

!HBONDS IN BETA SHEET

assi (resi   4 and name N ) (resi  52 and name O ) 3.3 0.8 0.0
assi (resi   4 and name HN ) (resi  52 and name O ) 2.3 0.5 0.0
assi (resi   6 and name N ) (resi  54 and name O ) 3.3 0.8 0.0
assi (resi   6 and name HN ) (resi  54 and name O ) 2.3 0.5 0.0
assi (resi   7 and name N ) (resi  77 and name O ) 3.3 0.8 0.0
assi (resi   7 and name HN ) (resi  77 and name O ) 2.3 0.5 0.0
assi (resi   8 and name N ) (resi  56 and name O ) 3.3 0.8 0.0
assi (resi   8 and name HN ) (resi  56 and name O ) 2.3 0.5 0.0
assi (resi   9 and name N ) (resi  79 and name O ) 3.3 0.8 0.0
assi (resi   9 and name HN ) (resi  79 and name O ) 2.3 0.5 0.0
assi (resi  42 and name N ) (resi  53 and name O ) 3.3 0.8 0.0
assi (resi  42 and name HN ) (resi  53 and name O ) 2.3 0.5 0.0
assi (resi  44 and name N ) (resi  51 and name O ) 3.3 0.8 0.0
assi (resi  44 and name HN ) (resi  51 and name O ) 2.3 0.5 0.0
assi (resi  46 and name N ) (resi  49 and name O ) 3.3 0.8 0.0
assi (resi  46 and name HN ) (resi  49 and name O ) 2.3 0.5 0.0
assi (resi  51 and name N ) (resi  44 and name O ) 3.3 0.8 0.0
assi (resi  51 and name HN ) (resi  44 and name O ) 2.3 0.5 0.0
assi (resi  53 and name N ) (resi  42 and name O ) 3.3 0.8 0.0
assi (resi  53 and name HN ) (resi  42 and name O ) 2.3 0.5 0.0
assi (resi  54 and name N ) (resi   4 and name O ) 3.3 0.8 0.0
assi (resi  54 and name HN ) (resi   4 and name O ) 2.3 0.5 0.0
assi (resi  55 and name N ) (resi  40 and name O ) 3.3 0.8 0.0
assi (resi  55 and name HN ) (resi  40 and name O ) 2.3 0.5 0.0
assi (resi  56 and name N ) (resi   6 and name O ) 3.3 0.8 0.0
assi (resi  56 and name HN ) (resi   6 and name O ) 2.3 0.5 0.0
assi (resi  77 and name N ) (resi   5 and name O ) 3.3 0.8 0.0
assi (resi  77 and name HN ) (resi   5 and name O ) 2.3 0.5 0.0
assi (resi  78 and name N ) (resi 109 and name O ) 3.3 0.8 0.0
assi (resi  78 and name HN ) (resi 109 and name O ) 2.3 0.5 0.0
assi (resi  79 and name N ) (resi   7 and name O ) 3.3 0.8 0.0
assi (resi  79 and name HN ) (resi   7 and name O ) 2.3 0.5 0.0
assi (resi  80 and name N ) (resi 111 and name O ) 3.3 0.8 0.0
assi (resi  80 and name HN ) (resi 111 and name O ) 2.3 0.5 0.0
assi (resi  81 and name N ) (resi   9 and name O ) 3.3 0.8 0.0
assi (resi  81 and name HN ) (resi   9 and name O ) 2.3 0.5 0.0
assi (resi  82 and name N ) (resi 113 and name O ) 3.3 0.8 0.0
assi (resi  82 and name HN ) (resi 113 and name O ) 2.3 0.5 0.0
assi (resi 111 and name N ) (resi  78 and name O ) 3.3 0.8 0.0
assi (resi 111 and name HN ) (resi  78 and name O ) 2.3 0.5 0.0
assi (resi 112 and name N ) (resi 153 and name O ) 3.3 0.8 0.0
assi (resi 112 and name HN ) (resi 153 and name O ) 2.3 0.5 0.0
assi (resi 113 and name N ) (resi  80 and name O ) 3.3 0.8 0.0
assi (resi 113 and name HN ) (resi  80 and name O ) 2.3 0.5 0.0
assi (resi 114 and name N ) (resi 155 and name O ) 3.3 0.8 0.0
assi (resi 114 and name HN ) (resi 155 and name O ) 2.3 0.5 0.0
assi (resi 153 and name N ) (resi 110 and name O ) 3.3 0.8 0.0
assi (resi 153 and name HN ) (resi 110 and name O ) 2.3 0.5 0.0
assi (resi 155 and name N ) (resi 112 and name O ) 3.3 0.8 0.0
assi (resi 155 and name HN ) (resi 112 and name O ) 2.3 0.5 0.0
```

Dihedral Angle Restraints

```
!! Phi restraints: a-helix  =  -80 +/- 50 degrees
!!                 b-strand = -105 +/- 65 degrees
!!
!! b-strand 4-8 (parallel)
 !!   4
 assign (resid   3 and name c)(resid   4 and name n)
  (resid   4 and name ca)(resid   4 and name c)  1.0 -105.0 65.0 2
 !!   5
 assign (resid   4 and name c)(resid   5 and name n)
  (resid   5 and name ca)(resid   5 and name c)  1.0 -105.0 65.0 2
 !!   6
 assign (resid   5 and name c)(resid   6 and name n)
  (resid   6 and name ca)(resid   6 and name c)  1.0 -105.0 65.0 2
 !!   7
 assign (resid   6 and name c)(resid   7 and name n)
  (resid   7 and name ca)(resid   7 and name c)  1.0 -105.0 65.0 2
 !!   8
 assign (resid   7 and name c)(resid   8 and name n)
  (resid   8 and name ca)(resid   8 and name c)  1.0 -105.0 65.0 2
!!
!! a-helix 16-24
 !!  16
 assign (resid  15 and name c)(resid  16 and name n)
  (resid  16 and name ca)(resid  16 and name c)  1.0  -80.0 50.0 2
 !!  17
 assign (resid  16 and name c)(resid  17 and name n)
  (resid  17 and name ca)(resid  17 and name c)  1.0  -80.0 50.0 2
 !!  18
 assign (resid  17 and name c)(resid  18 and name n)
  (resid  18 and name ca)(resid  18 and name c)  1.0  -80.0 50.0 2
 !!  19
 assign (resid  18 and name c)(resid  19 and name n)
  (resid  19 and name ca)(resid  19 and name c)  1.0  -80.0 50.0 2
 !!  20
 assign (resid  19 and name c)(resid  20 and name n)
  (resid  20 and name ca)(resid  20 and name c)  1.0  -80.0 50.0 2
 !!  21
 assign (resid  20 and name c)(resid  21 and name n)
  (resid  21 and name ca)(resid  21 and name c)  1.0  -80.0 50.0 2
 !!  22
 assign (resid  21 and name c)(resid  22 and name n)
  (resid  22 and name ca)(resid  22 and name c)  1.0  -80.0 50.0 2
 !!  23
 assign (resid  22 and name c)(resid  23 and name n)
  (resid  23 and name ca)(resid  23 and name c)  1.0  -80.0 50.0 2
 !!  23
 assign (resid  23 and name c)(resid  24 and name n)
  (resid  24 and name ca)(resid  24 and name c)  1.0  -80.0 50.0 2
!!
!! b-strand 41-44 (anti-parallel)
 !!  41
 assign (resid  40 and name c)(resid  41 and name n)
  (resid  41 and name ca)(resid  41 and name c)  1.0 -105.0 65.0 2
 !!  42
 assign (resid  41 and name c)(resid  42 and name n)
  (resid  42 and name ca)(resid  42 and name c)  1.0 -105.0 65.0 2

 !!  43
 assign (resid  42 and name c)(resid  43 and name n)
  (resid  43 and name ca)(resid  43 and name c)  1.0 -105.0 65.0 2
 !!  44
 assign (resid  43 and name c)(resid  44 and name n)
  (resid  44 and name ca)(resid  44 and name c)  1.0 -105.0 65.0 2
!!
!! b-strand 51-56 (both parallel and anti-parallel)
 !!  51
 assign (resid  50 and name c)(resid  51 and name n)
  (resid  51 and name ca)(resid  51 and name c)  1.0 -105.0 65.0 2
 !!  52
 assign (resid  51 and name c)(resid  52 and name n)
  (resid  52 and name ca)(resid  52 and name c)  1.0 -105.0 65.0 2
 !!  53
 assign (resid  52 and name c)(resid  53 and name n)
  (resid  53 and name ca)(resid  53 and name c)  1.0 -105.0 65.0 2
 !!  54
 assign (resid  53 and name c)(resid  54 and name n)
  (resid  54 and name ca)(resid  54 and name c)  1.0 -105.0 65.0 2
 !!  55
 assign (resid  54 and name c)(resid  55 and name n)
  (resid  55 and name ca)(resid  55 and name c)  1.0 -105.0 65.0 2
 !!  56
 assign (resid  55 and name c)(resid  56 and name n)
  (resid  56 and name ca)(resid  56 and name c)  1.0 -105.0 65.0 2
!!
!! b-strand 78-81 (parallel)
 !!  78
 assign (resid  77 and name c)(resid  78 and name n)
  (resid  78 and name ca)(resid  78 and name c)  1.0 -105.0 65.0 2
 !!  79
 assign (resid  78 and name c)(resid  79 and name n)
  (resid  79 and name ca)(resid  79 and name c)  1.0 -105.0 65.0 2
 !!  80
 assign (resid  79 and name c)(resid  80 and name n)
  (resid  80 and name ca)(resid  80 and name c)  1.0 -105.0 65.0 2
 !!  81
 assign (resid  80 and name c)(resid  81 and name n)
  (resid  81 and name ca)(resid  81 and name c)  1.0 -105.0 65.0 2
!!
!! a-helix 90-98
 !!  90
 assign (resid  89 and name c)(resid  90 and name n)
  (resid  90 and name ca)(resid  90 and name c)  1.0  -80.0 50.0 2
 !!  91
 assign (resid  90 and name c)(resid  91 and name n)
  (resid  91 and name ca)(resid  91 and name c)  1.0  -80.0 50.0 2
 !!  92
 assign (resid  91 and name c)(resid  92 and name n)
  (resid  92 and name ca)(resid  92 and name c)  1.0  -80.0 50.0 2
 !!  93
 assign (resid  92 and name c)(resid  93 and name n)
  (resid  93 and name ca)(resid  93 and name c)  1.0  -80.0 50.0 2
 !!  94
 assign (resid  93 and name c)(resid  94 and name n)
  (resid  94 and name ca)(resid  94 and name c)  1.0  -80.0 50.0 2
 !!  95
 assign (resid  94 and name c)(resid  95 and name n)
  (resid  95 and name ca)(resid  95 and name c)  1.0  -80.0 50.0 2
 !!  96
 assign (resid  95 and name c)(resid  96 and name n)
  (resid  96 and name ca)(resid  96 and name c)  1.0  -80.0 50.0 2
 !!  97
 assign (resid  96 and name c)(resid  97 and name n)
  (resid  97 and name ca)(resid  97 and name c)  1.0  -80.0 50.0 2
 !!  98
 assign (resid  97 and name c)(resid  98 and name n)
  (resid  98 and name ca)(resid  98 and name c)  1.0  -80.0 50.0 2
 !! 100 - Loose restraints (+/-70)
 assign (resid  99 and name c)(resid 100 and name n)
  (resid 100 and name ca)(resid 100 and name c)  1.0  -80.0 70.0 2
 !! 101 - Loose restraints (+/-70)
 assign (resid 100 and name c)(resid 101 and name n)
  (resid 101 and name ca)(resid 101 and name c)  1.0  -80.0 70.0 2
 !! 102 - Loose restraints (+/-70)
 assign (resid 101 and name c)(resid 102 and name n)
  (resid 102 and name ca)(resid 102 and name c)  1.0  -80.0 70.0 2
!!
!! b-strand 111-113 (parallel)
 !! 111
 assign (resid 110 and name c)(resid 111 and name n)
  (resid 111 and name ca)(resid 111 and name c)  1.0 -105.0 65.0 2
 !! 112
 assign (resid 111 and name c)(resid 112 and name n)
  (resid 112 and name ca)(resid 112 and name c)  1.0 -105.0 65.0 2
 !! 113
 assign (resid 112 and name c)(resid 113 and name n)
  (resid 113 and name ca)(resid 113 and name c)  1.0 -105.0 65.0 2
!!
!! a-helix 141-147
 !! 141
 assign (resid 140 and name c)(resid 141 and name n)
  (resid 141 and name ca)(resid 141 and name c)  1.0  -80.0 50.0 2
 !! 142
 assign (resid 141 and name c)(resid 142 and name n)
  (resid 142 and name ca)(resid 142 and name c)  1.0  -80.0 50.0 2
 !! 143
 assign (resid 142 and name c)(resid 143 and name n)
  (resid 143 and name ca)(resid 143 and name c)  1.0  -80.0 50.0 2
 !! 144
 assign (resid 143 and name c)(resid 144 and name n)
  (resid 144 and name ca)(resid 144 and name c)  1.0  -80.0 50.0 2
 !! 145
 assign (resid 144 and name c)(resid 145 and name n)
  (resid 145 and name ca)(resid 145 and name c)  1.0  -80.0 50.0 2
 !! 146
 assign (resid 145 and name c)(resid 146 and name n)
  (resid 146 and name ca)(resid 146 and name c)  1.0  -80.0 50.0 2
 !! 147
 assign (resid 146 and name c)(resid 147 and name n)
  (resid 147 and name ca)(resid 147 and name c)  1.0  -80.0 50.0 2
```

```
!!
!! b-strand 153-154 (parallel)
 !! 153
 assign (resid 152 and name c)(resid 153 and name n)
  (resid 153 and name ca)(resid 153 and name c)  1.0 -105.0 65.0 2
 !! 154
 assign (resid 153 and name c)(resid 154 and name n)
  (resid 154 and name ca)(resid 154 and name c)  1.0 -105.0 65.0 2
!!
!! a-helix 165-176
 !! 165
 assign (resid 164 and name c)(resid 165 and name n)
  (resid 165 and name ca)(resid 165 and name c)  1.0  -80.0 50.0 2
 !! 166
 assign (resid 165 and name c)(resid 166 and name n)
  (resid 166 and name ca)(resid 166 and name c)  1.0  -80.0 50.0 2
 !! 167
 assign (resid 166 and name c)(resid 167 and name n)
  (resid 167 and name ca)(resid 167 and name c)  1.0  -80.0 50.0 2
 !! 168
 assign (resid 167 and name c)(resid 168 and name n)
  (resid 168 and name ca)(resid 168 and name c)  1.0  -80.0 50.0 2
 !! 169
 assign (resid 168 and name c)(resid 169 and name n)
  (resid 169 and name ca)(resid 169 and name c)  1.0  -80.0 50.0 2
 !! 170
 assign (resid 169 and name c)(resid 170 and name n)
  (resid 170 and name ca)(resid 170 and name c)  1.0  -80.0 50.0 2
 !! 171
 assign (resid 170 and name c)(resid 171 and name n)
  (resid 171 and name ca)(resid 171 and name c)  1.0  -80.0 50.0 2
 !! 172
 assign (resid 171 and name c)(resid 172 and name n)
  (resid 172 and name ca)(resid 172 and name c)  1.0  -80.0 50.0 2
 !! 173
 assign (resid 172 and name c)(resid 173 and name n)
  (resid 173 and name ca)(resid 173 and name c)  1.0  -80.0 50.0 2
 !! 174
 assign (resid 173 and name c)(resid 174 and name n)
  (resid 174 and name ca)(resid 174 and name c)  1.0  -80.0 50.0 2
 !! 175
 assign (resid 174 and name c)(resid 175 and name n)
  (resid 175 and name ca)(resid 175 and name c)  1.0  -80.0 50.0 2
 !! 176
 assign (resid 175 and name c)(resid 176 and name n)
  (resid 176 and name ca)(resid 176 and name c)  1.0  -80.0 50.0 2
 !! 177
 assign (resid 176 and name c)(resid 177 and name n)
  (resid 177 and name ca)(resid 177 and name c)  1.0  -80.0 50.0 2
!!
!!
!! Psi restraints: a-helix  =  -20 +/- 50 degrees
!!                 b-strand =  145 +/- 45 degrees
!!
!! b-strand 4-8 (parallel)
 !!  4
 assign (resid  4 and name n)(resid  4 and name ca)
  (resid  4 and name c)(resid  5 and name n)  1.0  145.0 45.0 2
 !!  5
 assign (resid  5 and name n)(resid  5 and name ca)
  (resid  5 and name c)(resid  6 and name n)  1.0  145.0 45.0 2
 !!  6
 assign (resid  6 and name n)(resid  6 and name ca)
  (resid  6 and name c)(resid  7 and name n)  1.0  145.0 45.0 2
 !!  7
 assign (resid  7 and name n)(resid  7 and name ca)
  (resid  7 and name c)(resid  8 and name n)  1.0  145.0 45.0 2

 !!  8
 assign (resid  8 and name n)(resid  8 and name ca)
  (resid  8 and name c)(resid  9 and name n)  1.0  145.0 45.0 2
!!
!! a-helix 16-24
 !! 16
 assign (resid  16 and name n)(resid  16 and name ca)
  (resid  16 and name c)(resid  17 and name n)  1.0  -20.0 50.0 2
 !! 17
 assign (resid  17 and name n)(resid  17 and name ca)
  (resid  17 and name c)(resid  18 and name n)  1.0  -20.0 50.0 2
 !! 18
 assign (resid  18 and name n)(resid  18 and name ca)
  (resid  18 and name c)(resid  19 and name n)  1.0  -20.0 50.0 2
 !! 19
 assign (resid  19 and name n)(resid  19 and name ca)
  (resid  19 and name c)(resid  20 and name n)  1.0  -20.0 50.0 2
 !! 20
 assign (resid  20 and name n)(resid  20 and name ca)
  (resid  20 and name c)(resid  21 and name n)  1.0  -20.0 50.0 2
 !! 21
 assign (resid  21 and name n)(resid  21 and name ca)
  (resid  21 and name c)(resid  22 and name n)  1.0  -20.0 50.0 2
 !! 22
 assign (resid  22 and name n)(resid  22 and name ca)
  (resid  22 and name c)(resid  23 and name n)  1.0  -20.0 50.0 2
 !! 23
 assign (resid  23 and name n)(resid  23 and name ca)
  (resid  23 and name c)(resid  24 and name n)  1.0  -20.0 50.0 2
 !! 24
 assign (resid  24 and name n)(resid  24 and name ca)
  (resid  24 and name c)(resid  25 and name n)  1.0  -20.0 50.0 2
!!
!! b-strand 41-44 (anti-parallel)
 !! 41
 assign (resid  41 and name n)(resid  41 and name ca)
  (resid  41 and name c)(resid  42 and name n)  1.0  145.0 45.0 2
 !! 42
 assign (resid  42 and name n)(resid  42 and name ca)
  (resid  42 and name c)(resid  43 and name n)  1.0  145.0 45.0 2
 !! 43
 assign (resid  43 and name n)(resid  43 and name ca)
  (resid  43 and name c)(resid  44 and name n)  1.0  145.0 45.0 2
 !! 44
 assign (resid  44 and name n)(resid  44 and name ca)
  (resid  44 and name c)(resid  45 and name n)  1.0  145.0 45.0 2
!!
!! b-strand 51-56 (both parallel and anti-parallel)
 !! 51
 assign (resid  51 and name n)(resid  51 and name ca)
  (resid  51 and name c)(resid  52 and name n)  1.0  145.0 45.0 2
 !! 52
 assign (resid  52 and name n)(resid  52 and name ca)
  (resid  52 and name c)(resid  53 and name n)  1.0  145.0 45.0 2
 !! 53
 assign (resid  53 and name n)(resid  53 and name ca)
  (resid  53 and name c)(resid  54 and name n)  1.0  145.0 45.0 2
 !! 54
 assign (resid  54 and name n)(resid  54 and name ca)
  (resid  54 and name c)(resid  55 and name n)  1.0  145.0 45.0 2
 !! 55
 assign (resid  55 and name n)(resid  55 and name ca)
  (resid  55 and name c)(resid  56 and name n)  1.0  145.0 45.0 2
 !! 56
 assign (resid  56 and name n)(resid  56 and name ca)
  (resid  56 and name c)(resid  57 and name n)  1.0  145.0 45.0 2
!!
!! b-strand 78-81 (parallel)
 !! 78
 assign (resid  78 and name n)(resid  78 and name ca)
  (resid  78 and name c)(resid  79 and name n)  1.0  145.0 45.0 2
 !! 79
 assign (resid  79 and name n)(resid  79 and name ca)
  (resid  79 and name c)(resid  80 and name n)  1.0  145.0 45.0 2
 !! 80
 assign (resid  80 and name n)(resid  80 and name ca)
  (resid  80 and name c)(resid  81 and name n)  1.0  145.0 45.0 2
 !! 81
 assign (resid  81 and name n)(resid  81 and name ca)
  (resid  81 and name c)(resid  82 and name n)  1.0  145.0 45.0 2
!!
!! a-helix 90-98
 !! 90
 assign (resid  90 and name n)(resid  90 and name ca)
  (resid  90 and name c)(resid  91 and name n)  1.0  -20.0 50.0 2
 !! 91
 assign (resid  91 and name n)(resid  91 and name ca)
  (resid  91 and name c)(resid  92 and name n)  1.0  -20.0 50.0 2
 !! 92
 assign (resid  92 and name n)(resid  92 and name ca)
  (resid  92 and name c)(resid  93 and name n)  1.0  -20.0 50.0 2
 !! 93
 assign (resid  93 and name n)(resid  93 and name ca)
  (resid  93 and name c)(resid  94 and name n)  1.0  -20.0 50.0 2
 !! 94
 assign (resid  94 and name n)(resid  94 and name ca)
  (resid  94 and name c)(resid  95 and name n)  1.0  -20.0 50.0 2
 !! 95
 assign (resid  95 and name n)(resid  95 and name ca)
  (resid  95 and name c)(resid  96 and name n)  1.0  -20.0 50.0 2
 !! 96
 assign (resid  96 and name n)(resid  96 and name ca)
  (resid  96 and name c)(resid  97 and name n)  1.0  -20.0 50.0 2
 !! 97
 assign (resid  97 and name n)(resid  97 and name ca)
  (resid  97 and name c)(resid  98 and name n)  1.0  -20.0 50.0 2
 !! 98
 assign (resid  98 and name n)(resid  98 and name ca)
  (resid  98 and name c)(resid  99 and name n)  1.0  -20.0 50.0 2
```

```
!! 100 - Loose (+/-70)
 assign (resid 100 and name n)(resid 100 and name ca)
  (resid 100 and name c)(resid 101 and name n)  1.0   -20.0 70.0 2
  !! 101 - Loose (+/-70)
 assign (resid 101 and name n)(resid 101 and name ca)
  (resid 101 and name c)(resid 102 and name n)  1.0   -20.0 70.0 2
  !! 100 - Loose (+/-70)
 assign (resid 102 and name n)(resid 102 and name ca)
  (resid 102 and name c)(resid 103 and name n)  1.0   -20.0 70.0 2


!!
!! b-strand 111-113 (parallel)
  !! 111
  assign (resid 111 and name n)(resid 111 and name ca)
  (resid 111 and name c)(resid 112 and name n)  1.0   145.0 45.0 2
  !! 112
  assign (resid 112 and name n)(resid 112 and name ca)
  (resid 112 and name c)(resid 113 and name n)  1.0   145.0 45.0 2
  !! 113
  assign (resid 113 and name n)(resid 113 and name ca)
  (resid 113 and name c)(resid 114 and name n)  1.0   145.0 45.0 2
!!
!! a-helix 141-147
  !! 141
 assign (resid 141 and name n)(resid 141 and name ca)
  (resid 141 and name c)(resid 142 and name n)  1.0   -20.0 50.0 2
  !! 142
 assign (resid 142 and name n)(resid 142 and name ca)
  (resid 142 and name c)(resid 143 and name n)  1.0   -20.0 50.0 2
  !! 143
 assign (resid 143 and name n)(resid 143 and name ca)
  (resid 143 and name c)(resid 144 and name n)  1.0   -20.0 50.0 2
  !! 144
 assign (resid 144 and name n)(resid 144 and name ca)
  (resid 144 and name c)(resid 145 and name n)  1.0   -20.0 50.0 2
  !! 145
 assign (resid 145 and name n)(resid 145 and name ca)
  (resid 145 and name c)(resid 146 and name n)  1.0   -20.0 50.0 2
  !! 146
 assign (resid 146 and name n)(resid 146 and name ca)
  (resid 146 and name c)(resid 147 and name n)  1.0   -20.0 50.0 2
  !! 147
 assign (resid 147 and name n)(resid 147 and name ca)
  (resid 147 and name c)(resid 148 and name n)  1.0   -20.0 50.0 2
!!
!! b-strand 153-154 (parallel)
  !! 153
  assign (resid 153 and name n)(resid 153 and name ca)
  (resid 153 and name c)(resid 154 and name n)  1.0   145.0 45.0 2
  !! 154
  assign (resid 154 and name n)(resid 154 and name ca)
  (resid 154 and name c)(resid 155 and name n)  1.0   145.0 45.0 2
!!
!! a-helix 165-176
  !! 165
 assign (resid 165 and name n)(resid 165 and name ca)
  (resid 165 and name c)(resid 166 and name n)  1.0   -20.0 50.0 2
  !! 166
 assign (resid 166 and name n)(resid 166 and name ca)
  (resid 166 and name c)(resid 167 and name n)  1.0   -20.0 50.0 2
  !! 167
 assign (resid 167 and name n)(resid 167 and name ca)
  (resid 167 and name c)(resid 168 and name n)  1.0   -20.0 50.0 2
  !! 168
 assign (resid 168 and name n)(resid 168 and name ca)
  (resid 168 and name c)(resid 169 and name n)  1.0   -20.0 50.0 2
  !! 169
 assign (resid 169 and name n)(resid 169 and name ca)
  (resid 169 and name c)(resid 170 and name n)  1.0   -20.0 50.0 2

  !! 170
 assign (resid 170 and name n)(resid 170 and name ca)
  (resid 170 and name c)(resid 171 and name n)  1.0   -20.0 50.0 2
  !! 171
 assign (resid 171 and name n)(resid 171 and name ca)
  (resid 171 and name c)(resid 172 and name n)  1.0   -20.0 50.0 2
  !! 172
 assign (resid 172 and name n)(resid 172 and name ca)
  (resid 172 and name c)(resid 173 and name n)  1.0   -20.0 50.0 2
  !! 173
 assign (resid 173 and name n)(resid 173 and name ca)
  (resid 173 and name c)(resid 174 and name n)  1.0   -20.0 50.0 2
  !! 174
 assign (resid 174 and name n)(resid 174 and name ca)
  (resid 174 and name c)(resid 175 and name n)  1.0   -20.0 50.0 2
  !! 175
 assign (resid 175 and name n)(resid 175 and name ca)
  (resid 175 and name c)(resid 176 and name n)  1.0   -20.0 50.0 2
  !! 176
 assign (resid 176 and name n)(resid 176 and name ca)
  (resid 176 and name c)(resid 177 and name n)  1.0   -20.0 50.0 2
  !! 177
 assign (resid 177 and name n)(resid 177 and name ca)
  (resid 177 and name c)(resid 178 and name n)  1.0   -20.0 50.0 2
```

Hα Chemical Shift Restraints

```
!! GLY -7
 !!OBSE (resid -7 and (name HA1)) 3.97
 !! SER -6
 !OBSE (resid -6 and (name HA)) 4.50
 !! LYS -5
 !OBSE (resid -5 and (name HA)) 4.44
 !! SER -6
 !OBSE (resid -6 and (name HA)) 4.56
 !! LYS -5
 !OBSE (resid -5 and (name HA)) 4.43
 !! ILE -4
OBSE (resid -4 and (name HA)) 4.22
 !! ILE -3
OBSE (resid -3 and (name HA)) 4.28
 !! SER -2
OBSE (resid -2 and (name HA)) 4.51
 !! ALA -1
OBSE (resid -1 and (name HA)) 4.37
 !! MET 1
OBSE (resid 1 and (name HA)) 4.46
 !! GLN 2
OBSE (resid 2 and (name HA)) 4.45
 !! THR 3
OBSE (resid 3 and (name HA)) 4.45
 !! ILE 4
OBSE (resid 4 and (name HA)) 4.54
 !! LYS 5
OBSE (resid 5 and (name HA)) 4.92
 !! CYS 6
OBSE (resid 6 and (name HA)) 5.72
 !! VAL 7
OBSE (resid 7 and (name HA)) 5.10
 !! VAL 8
OBSE (resid 8 and (name HA)) 4.90
 !! VAL 9
OBSE (resid 9 and (name HA)) 4.49
 !! ASP 11
OBSE (resid 11 and (name HA)) 3.96
 !! GLY 12
OBSE (resid 12 and (name HA1)) 4.14
 !! GLY 12
OBSE (resid 12 and (name HA2)) 3.78
 !! ALA 13
OBSE (resid 13 and (name HA)) 4.17
 !! VAL 14
OBSE (resid 14 and (name HA)) 4.46
 !! GLY 15
OBSE (resid 15 and (name HA1)) 4.92
 !! GLY 15
OBSE (resid 15 and (name HA2)) 4.41
 !! LYS 16
OBSE (resid 16 and (name HA)) 3.68
 !! THR 17
OBSE (resid 17 and (name HA)) 4.03
 !! CYS 18
OBSE (resid 18 and (name HA)) 3.69
 !! LEU 19
OBSE (resid 19 and (name HA)) 4.06
 !! LEU 20
OBSE (resid 20 and (name HA)) 3.91
 !! ILE 21
OBSE (resid 21 and (name HA)) 3.20
 !! SER 22
OBSE (resid 22 and (name HA)) 4.21
 !! THR 24
OBSE (resid 24 and (name HA)) 3.71
 !! THR 25
OBSE (resid 25 and (name HA)) 4.64
 !! ASN 26
OBSE (resid 26 and (name HA)) 4.61
 !! LYS 27
OBSE (resid 27 and (name HA)) 4.54
 !! PHE 28
OBSE (resid 28 and (name HA)) 4.56
 !! PRO 29
OBSE (resid 29 and (name HA)) 4.53
 !! SER 30
OBSE (resid 30 and (name HA)) 4.32
 !! GLU 31
OBSE (resid 31 and (name HA)) 4.41
 !! TYR 32
OBSE (resid 32 and (name HA)) 4.48
 !! VAL 33
OBSE (resid 33 and (name HA)) 4.19
 !! THR 35
OBSE (resid 35 and (name HA)) 4.77
 !! TYR 40
OBSE (resid 40 and (name HA)) 4.66
 !! ALA 41
OBSE (resid 41 and (name HA)) 5.67
 !! VAL 42
OBSE (resid 42 and (name HA)) 4.67
 !! THR 43
OBSE (resid 43 and (name HA)) 5.23
 !! VAL 44
OBSE (resid 44 and (name HA)) 4.50
 !! MET 45
OBSE (resid 45 and (name HA)) 5.37
 !! ILE 46
OBSE (resid 46 and (name HA)) 4.48
 !! GLY 47
OBSE (resid 47 and (name HA1)) 4.82
 !! GLY 47
OBSE (resid 47 and (name HA2)) 3.96
 !! GLY 48
OBSE (resid 48 and (name HA1)) 4.23
 !! GLY 48
OBSE (resid 48 and (name HA2)) 3.70
 !! GLU 49
OBSE (resid 49 and (name HA)) 5.04
 !! PRO 50
OBSE (resid 50 and (name HA)) 5.16
 !! TYR 51
OBSE (resid 51 and (name HA)) 4.80
 !! THR 52
OBSE (resid 52 and (name HA)) 4.84
 !! LEU 53
OBSE (resid 53 and (name HA)) 5.15
 !! GLY 54
OBSE (resid 54 and (name HA1)) 4.72
 !! GLY 54
OBSE (resid 54 and (name HA2)) 3.70
 !! LEU 55
OBSE (resid 55 and (name HA)) 5.22
 !! PHE 56
OBSE (resid 56 and (name HA)) 4.86
 !! THR 58
OBSE (resid 58 and (name HA)) 3.60
 !! GLN 61
OBSE (resid 61 and (name HA)) 5.06
 !! GLU 62
OBSE (resid 62 and (name HA)) 4.34
 !! ASP 63
OBSE (resid 63 and (name HA)) 4.51
 !! TYR 64
OBSE (resid 64 and (name HA)) 4.62
 !! ASP 65
OBSE (resid 65 and (name HA)) 4.62
 !! LEU 70
OBSE (resid 70 and (name HA)) 4.14
 !! SER 71
OBSE (resid 71 and (name HA)) 4.55
 !! TYR 72
OBSE (resid 72 and (name HA)) 4.07
 !! PRO 73
OBSE (resid 73 and (name HA)) 4.54
 !! GLN 74
OBSE (resid 74 and (name HA)) 4.49
 !! ASP 76
OBSE (resid 76 and (name HA)) 5.08
 !! VAL 77
OBSE (resid 77 and (name HA)) 4.81
 !! PHE 78
OBSE (resid 78 and (name HA)) 5.71
 !! LEU 79
OBSE (resid 79 and (name HA)) 4.79
 !! VAL 80
OBSE (resid 80 and (name HA)) 4.27
 !! CYS 81
OBSE (resid 81 and (name HA)) 5.86
 !! PHE 82
OBSE (resid 82 and (name HA)) 4.61
 !! SER 83
OBSE (resid 83 and (name HA)) 5.30
 !! VAL 84
OBSE (resid 84 and (name HA)) 4.36
 !! VAL 85
OBSE (resid 85 and (name HA)) 4.57
 !! SER 86
OBSE (resid 86 and (name HA)) 5.27
```

```
 !! SER 88
OBSE (resid 88 and (name HA)) 4.83
 !! SER 89
OBSE (resid 89 and (name HA)) 5.14
 !! PHE 90
OBSE (resid 90 and (name HA)) 4.03
 !! GLU 91
OBSE (resid 91 and (name HA)) 4.03
 !! ASN 92
OBSE (resid 92 and (name HA)) 4.86
 !! VAL 93
OBSE (resid 93 and (name HA)) 3.86
 !! LYS 94
OBSE (resid 94 and (name HA)) 4.01
 !! GLU 95
OBSE (resid 95 and (name HA)) 4.17
 !! LYS 96
OBSE (resid 96 and (name HA)) 4.37
 !! TRP 97
OBSE (resid 97 and (name HA)) 4.52
 !! VAL 98
OBSE (resid 98 and (name HA)) 3.63
 !! GLU 100
OBSE (resid 100 and (name HA)) 4.27
 !! ILE 101
OBSE (resid 101 and (name HA)) 3.62
 !! THR 102
OBSE (resid 102 and (name HA)) 3.92
 !! HIS 103
OBSE (resid 103 and (name HA)) 4.42
 !! HIS 104
OBSE (resid 104 and (name HA)) 4.28
 !! CYS 105
OBSE (resid 105 and (name HA)) 5.53
 !! PRO 106
OBSE (resid 106 and (name HA)) 4.62
 !! LYS 107
OBSE (resid 107 and (name HA)) 4.70
 !! THR 108
OBSE (resid 108 and (name HA)) 4.83
 !! PRO 109
OBSE (resid 109 and (name HA)) 4.68
 !! PHE 110
OBSE (resid 110 and (name HA)) 6.01
 !! LEU 111
OBSE (resid 111 and (name HA)) 4.80
 !! LEU 112
OBSE (resid 112 and (name HA)) 5.46
 !! VAL 113
OBSE (resid 113 and (name HA)) 4.99
 !! GLY 114
OBSE (resid 114 and (name HA1)) 3.76
 !! GLN 116
OBSE (resid 116 and (name HA)) 4.09
 !! ILE 117
OBSE (resid 117 and (name HA)) 4.54
 !! ASP 118
OBSE (resid 118 and (name HA)) 4.58
 !! LEU 119
OBSE (resid 119 and (name HA)) 4.42
 !! ARG 120
OBSE (resid 120 and (name HA)) 3.68
 !! ASP 121
OBSE (resid 121 and (name HA)) 4.91
 !! ASP 122
OBSE (resid 122 and (name HA)) 5.07
 !! PRO 123
OBSE (resid 123 and (name HA)) 4.24
 !! SER 124
OBSE (resid 124 and (name HA)) 4.36
 !! THR 125
OBSE (resid 125 and (name HA)) 3.97
 !! ILE 126
OBSE (resid 126 and (name HA)) 3.63
 !! GLU 127
OBSE (resid 127 and (name HA)) 4.23
 !! LYS 128
OBSE (resid 128 and (name HA)) 4.09
 !! LEU 129
OBSE (resid 129 and (name HA)) 4.21
 !! ALA 130
OBSE (resid 130 and (name HA)) 4.39
 !! LYS 131
OBSE (resid 131 and (name HA)) 4.21
 !! ASN 132
OBSE (resid 132 and (name HA)) 5.09
 !! LYS 133
OBSE (resid 133 and (name HA)) 4.05
 !! GLN 134
OBSE (resid 134 and (name HA)) 4.87
 !! LYS 135
OBSE (resid 135 and (name HA)) 4.87
 !! PRO 136
OBSE (resid 136 and (name HA)) 4.50
 !! ILE 137
OBSE (resid 137 and (name HA)) 4.01
 !! THR 138
OBSE (resid 138 and (name HA)) 4.70
 !! PRO 139
OBSE (resid 139 and (name HA)) 4.18
 !! GLU 140
OBSE (resid 140 and (name HA)) 4.08
 !! THR 141
OBSE (resid 141 and (name HA)) 3.92
 !! ALA 142
OBSE (resid 142 and (name HA)) 3.96
 !! GLU 143
OBSE (resid 143 and (name HA)) 3.84
 !! LYS 144
OBSE (resid 144 and (name HA)) 4.04
 !! LEU 145
OBSE (resid 145 and (name HA)) 4.24
 !! ALA 146
OBSE (resid 146 and (name HA)) 3.82
 !! ARG 147
OBSE (resid 147 and (name HA)) 4.26
 !! ASP 148
OBSE (resid 148 and (name HA)) 4.43
 !! LEU 149
OBSE (resid 149 and (name HA)) 4.46
 !! LYS 150
OBSE (resid 150 and (name HA)) 3.87
 !! ALA 151
OBSE (resid 151 and (name HA)) 3.07
 !! VAL 152
OBSE (resid 152 and (name HA)) 3.59
 !! LYS 153
OBSE (resid 153 and (name HA)) 4.43
 !! TYR 154
OBSE (resid 154 and (name HA)) 5.85
 !! VAL 155
OBSE (resid 155 and (name HA)) 4.23
 !! GLU 156
OBSE (resid 156 and (name HA)) 5.48
 !! CYS 157
OBSE (resid 157 and (name HA)) 5.42
 !! SER 158
OBSE (resid 158 and (name HA)) 5.29
 !! ALA 159
OBSE (resid 159 and (name HA)) 4.30
 !! LEU 160
OBSE (resid 160 and (name HA)) 3.38
 !! THR 161
OBSE (resid 161 and (name HA)) 4.33
 !! GLN 162
OBSE (resid 162 and (name HA)) 4.04
 !! LYS 163
OBSE (resid 163 and (name HA)) 4.26
 !! GLY 164
OBSE (resid 164 and (name HA1)) 4.38
 !! GLY 164
OBSE (resid 164 and (name HA2)) 3.94
 !! LEU 165
OBSE (resid 165 and (name HA)) 3.88
 !! LYS 166
OBSE (resid 166 and (name HA)) 4.23
 !! ASN 167
OBSE (resid 167 and (name HA)) 4.42
 !! VAL 168
OBSE (resid 168 and (name HA)) 2.99
 !! PHE 169
OBSE (resid 169 and (name HA)) 3.84
 !! ASP 170
OBSE (resid 170 and (name HA)) 4.36
 !! GLU 171
OBSE (resid 171 and (name HA)) 4.09
 !! ALA 172
OBSE (resid 172 and (name HA)) 3.87
 !! ILE 173
OBSE (resid 173 and (name HA)) 3.43
 !! LEU 174
OBSE (resid 174 and (name HA)) 3.87
 !! ALA 175
OBSE (resid 175 and (name HA)) 4.13
 !! ALA 176
OBSE (resid 176 and (name HA)) 4.11
 !! LEU 177
OBSE (resid 177 and (name HA)) 4.29
```

```
 !! GLU 178
OBSE (resid 178 and (name HA))  4.67
 !! PRO 179
OBSE (resid 179 and (name HA))  4.83
 !! PRO 180
OBSE (resid 180 and (name HA))  4.54
 !! GLU 181
OBSE (resid 181 and (name HA))  4.67
 !! PRO 182
OBSE (resid 182 and (name HA))  4.48
```

```
 !! LYS 183
OBSE (resid 183 and (name HA))  4.36
 !! LYS 184
OBSE (resid 184 and (name HA))  4.45
 !! SER 185
OBSE (resid 185 and (name HA))  4.53
 !! ARG 186
OBSE (resid 186 and (name HA))  4.44
 !! ARG 187
OBSE (resid 187 and (name HA))  4.25
```

Carbon Chemical Shift Restraints

```
!! GLY -7
!!assign                  (resid -7 and name n )
!!     (resid -7 and name ca) (resid -7 and name c )
!!     (resid -6 and name n )                    41.59  0.0
 !! ILE -4
assign (resid -5 and name c ) (resid -4 and name n )
       (resid -4 and name ca) (resid -4 and name c )
       (resid -3 and name n )                  59.29  36.77
 !! ILE -3
assign (resid -4 and name c ) (resid -3 and name n )
       (resid -3 and name ca) (resid -3 and name c )
       (resid -2 and name n )                  59.25  36.86
 !! SER -2
assign (resid -3 and name c ) (resid -2 and name n )
       (resid -2 and name ca) (resid -2 and name c )
       (resid -1 and name n )                  56.43  62.26
 !! ALA -1
assign (resid -2 and name c ) (resid -1 and name n )
       (resid -1 and name ca) (resid -1 and name c )
       (resid 1 and name n )                  50.90  17.50
 !! MET 1
assign (resid -1 and name c ) (resid 1 and name n )
       (resid 1 and name ca) (resid 1 and name c )
       (resid 2 and name n )                  54.04  31.44
 !! GLN 2
assign (resid 1 and name c ) (resid 2 and name n )
       (resid 2 and name ca) (resid 2 and name c )
       (resid 3 and name n )                  53.71  28.35
 !! THR 3
assign (resid 2 and name c ) (resid 3 and name n )
       (resid 3 and name ca) (resid 3 and name c )
       (resid 4 and name n )                  59.87  68.68
 !! ILE 4
assign (resid 3 and name c ) (resid 4 and name n )
       (resid 4 and name ca) (resid 4 and name c )
       (resid 5 and name n )                  58.03  38.55
 !! LYS 5
assign (resid 4 and name c ) (resid 5 and name n )
       (resid 5 and name ca) (resid 5 and name c )
       (resid 6 and name n )                  54.51  1.0E+18
 !! CYS 6
assign (resid 5 and name c ) (resid 6 and name n )
       (resid 6 and name ca) (resid 6 and name c )
       (resid 7 and name n )                  54.49  26.18
 !! VAL 7
assign (resid 6 and name c ) (resid 7 and name n )
       (resid 7 and name ca) (resid 7 and name c )
       (resid 8 and name n )                  59.09  31.62
 !! VAL 8
assign (resid 7 and name c ) (resid 8 and name n )
       (resid 8 and name ca) (resid 8 and name c )
       (resid 9 and name n )                  59.79  30.88
 !! VAL 9
assign (resid 8 and name c ) (resid 9 and name n )
       (resid 9 and name ca) (resid 9 and name c )
       (resid 10 and name n )                  56.66  41.44

 !! GLY 10
assign (resid 9 and name c ) (resid 10 and name n )
       (resid 10 and name ca) (resid 10 and name c )
       (resid 11 and name n )                  56.27  0.0
 !! ASP 11
assign (resid 10 and name c ) (resid 11 and name n )
       (resid 11 and name ca) (resid 11 and name c )
       (resid 12 and name n )                  53.68  38.48
 !! GLY 12
assign (resid 11 and name c ) (resid 12 and name n )
       (resid 12 and name ca) (resid 12 and name c )
       (resid 13 and name n )                  45.31  0.0
 !! ALA 13
assign (resid 12 and name c ) (resid 13 and name n )
       (resid 13 and name ca) (resid 13 and name c )
       (resid 14 and name n )                  51.77  14.68
 !! VAL 14
assign (resid 13 and name c ) (resid 14 and name n )
       (resid 14 and name ca) (resid 14 and name c )
       (resid 15 and name n )                  60.37  1.0E+18
 !! GLY 15
assign (resid 14 and name c ) (resid 15 and name n )
       (resid 15 and name ca) (resid 15 and name c )
       (resid 16 and name n )                  42.99  0.0
 !! LYS 16
assign (resid 15 and name c ) (resid 16 and name n )
       (resid 16 and name ca) (resid 16 and name c )
       (resid 17 and name n )                  59.05  28.13
```

```
 !! THR 17
assign (resid 16 and name c ) (resid 17 and name n )
       (resid 17 and name ca) (resid 17 and name c )
       (resid 18 and name n )                  65.08  1.0E+18
 !! CYS 18
assign (resid 17 and name c ) (resid 18 and name n )
       (resid 18 and name ca) (resid 18 and name c )
       (resid 19 and name n )                  64.12  26.73
 !! LEU 19
assign (resid 18 and name c ) (resid 19 and name n )
       (resid 19 and name ca) (resid 19 and name c )
       (resid 20 and name n )                  57.71  39.91
 !! LEU 20
assign (resid 19 and name c ) (resid 20 and name n )
       (resid 20 and name ca) (resid 20 and name c )
       (resid 21 and name n )                  56.07  1.0E+18
 !! ILE 21
assign (resid 20 and name c ) (resid 21 and name n )
       (resid 21 and name ca) (resid 21 and name c )
       (resid 22 and name n )                  64.52  37.11
 !! SER 22
assign (resid 21 and name c ) (resid 22 and name n )
       (resid 22 and name ca) (resid 22 and name c )
       (resid 23 and name n )                  59.76  62.20
 !! THR 24
assign (resid 23 and name c ) (resid 24 and name n )
       (resid 24 and name ca) (resid 24 and name c )
       (resid 25 and name n )                  62.68  68.04
 !! THR 25
assign (resid 24 and name c ) (resid 25 and name n )
       (resid 25 and name ca) (resid 25 and name c )
       (resid 26 and name n )                  60.41  69.97
 !! ASN 26
assign (resid 25 and name c ) (resid 26 and name n )
       (resid 26 and name ca) (resid 26 and name c )
       (resid 27 and name n )                  53.58  36.50
 !! LYS 27
assign (resid 26 and name c ) (resid 27 and name n )
       (resid 27 and name ca) (resid 27 and name c )
       (resid 28 and name n )                  53.39  33.39
 !! PHE 28
assign (resid 27 and name c ) (resid 28 and name n )
       (resid 28 and name ca) (resid 28 and name c )
       (resid 29 and name n )                  53.69  1.0E+18
 !! PRO 29
assign (resid 28 and name c ) (resid 29 and name n )
       (resid 29 and name ca) (resid 29 and name c )
       (resid 30 and name n )                  60.41  28.34
 !! SER 30
assign (resid 29 and name c ) (resid 30 and name n )
       (resid 30 and name ca) (resid 30 and name c )
       (resid 31 and name n )                  57.32  62.17
 !! GLU 31
assign (resid 30 and name c ) (resid 31 and name n )
       (resid 31 and name ca) (resid 31 and name c )
       (resid 32 and name n )                  55.06  28.79
 !! TYR 32
assign (resid 31 and name c ) (resid 32 and name n )
       (resid 32 and name ca) (resid 32 and name c )
       (resid 33 and name n )                  56.48  37.18
 !! VAL 33
assign (resid 32 and name c ) (resid 33 and name n )
       (resid 33 and name ca) (resid 33 and name c )
       (resid 34 and name n )                  57.07  31.96
 !! THR 35
assign (resid 34 and name c ) (resid 35 and name n )
       (resid 35 and name ca) (resid 35 and name c )
       (resid 36 and name n )                  58.60  1.0E+18
 !! TYR 40
assign (resid 39 and name c ) (resid 40 and name n )
       (resid 40 and name ca) (resid 40 and name c )
       (resid 41 and name n )                  52.55  39.52
 !! ALA 41
assign (resid 40 and name c ) (resid 41 and name n )
       (resid 41 and name ca) (resid 41 and name c )
       (resid 42 and name n )                  49.06  19.13
 !! VAL 42
assign (resid 41 and name c ) (resid 42 and name n )
       (resid 42 and name ca) (resid 42 and name c )
       (resid 43 and name n )                  57.76  34.00
 !! THR 43
assign (resid 42 and name c ) (resid 43 and name n )
       (resid 43 and name ca) (resid 43 and name c )
       (resid 44 and name n )                  60.85  67.50
 !! VAL 44
assign (resid 43 and name c ) (resid 44 and name n )
      (resid 44 and name ca) (resid 44 and name c )
      (resid 45 and name n )                58.80  1.0E+18
 !! MET 45
assign (resid 44 and name c ) (resid 45 and name n )
      (resid 45 and name ca) (resid 45 and name c )
      (resid 46 and name n )                51.60  30.16
 !! ILE 46
assign (resid 45 and name c ) (resid 46 and name n )
      (resid 46 and name ca) (resid 46 and name c )
      (resid 47 and name n )                58.06  33.80
 !! GLY 47
assign (resid 46 and name c ) (resid 47 and name n )
      (resid 47 and name ca) (resid 47 and name c )
      (resid 48 and name n )                45.32  0.0
 !! GLY 48
assign (resid 47 and name c ) (resid 48 and name n )
      (resid 48 and name ca) (resid 48 and name c )
      (resid 49 and name n )                43.12  0.0
 !! GLU 49
assign (resid 48 and name c ) (resid 49 and name n )
      (resid 49 and name ca) (resid 49 and name c )
      (resid 50 and name n )                50.64  1.0E+18
 !! PRO 50
assign (resid 49 and name c ) (resid 50 and name n )
      (resid 50 and name ca) (resid 50 and name c )
      (resid 51 and name n )                60.34  30.39
 !! TYR 51
assign (resid 50 and name c ) (resid 51 and name n )
      (resid 51 and name ca) (resid 51 and name c )
      (resid 52 and name n )                55.75  40.40
 !! THR 52
assign (resid 51 and name c ) (resid 52 and name n )
      (resid 52 and name ca) (resid 52 and name c )
      (resid 53 and name n )                60.74  68.08
 !! LEU 53
assign (resid 52 and name c ) (resid 53 and name n )
      (resid 53 and name ca) (resid 53 and name c )
      (resid 54 and name n )                51.72  42.62
 !! GLY 54
assign (resid 53 and name c ) (resid 54 and name n )
      (resid 54 and name ca) (resid 54 and name c )
      (resid 55 and name n )                43.40  0.0
 !! LEU 55
assign (resid 54 and name c ) (resid 55 and name n )
      (resid 55 and name ca) (resid 55 and name c )
      (resid 56 and name n )                52.05  37.88
 !! PHE 56
assign (resid 55 and name c ) (resid 56 and name n )
      (resid 56 and name ca) (resid 56 and name c )
      (resid 57 and name n )                60.36  1.0E+18
 !! ASP 57
assign (resid 56 and name c ) (resid 57 and name n )
      (resid 57 and name ca) (resid 57 and name c )
      (resid 58 and name n )                55.95  1.0E+18
 !! THR 58
assign (resid 57 and name c ) (resid 58 and name n )
      (resid 58 and name ca) (resid 58 and name c )
      (resid 59 and name n )                56.90  1.0E+18
 !! ALA 59
assign (resid 58 and name c ) (resid 59 and name n )
      (resid 59 and name ca) (resid 59 and name c )
      (resid 60 and name n )                56.64  1.0E+18
 !! GLN 61
assign (resid 60 and name c ) (resid 61 and name n )
      (resid 61 and name ca) (resid 61 and name c )
      (resid 62 and name n )                60.15  1.0E+18
 !! GLU 62
assign (resid 61 and name c ) (resid 62 and name n )
      (resid 62 and name ca) (resid 62 and name c )
      (resid 63 and name n )                56.69  27.44
 !! ASP 63
assign (resid 62 and name c ) (resid 63 and name n )
      (resid 63 and name ca) (resid 63 and name c )
      (resid 64 and name n )                53.93  38.11
 !! TYR 64
assign (resid 63 and name c ) (resid 64 and name n )
      (resid 64 and name ca) (resid 64 and name c )
      (resid 65 and name n )                57.06  36.65
 !! LEU 70
assign (resid 69 and name c ) (resid 70 and name n )
      (resid 70 and name ca) (resid 70 and name c )
      (resid 71 and name n )                55.18  38.55
 !! SER 71
assign (resid 70 and name c ) (resid 71 and name n )
      (resid 71 and name ca) (resid 71 and name c )
      (resid 72 and name n )                58.00  1.0E+18
 !! TYR 72
assign (resid 71 and name c ) (resid 72 and name n )
      (resid 72 and name ca) (resid 72 and name c )
      (resid 73 and name n )                54.51  1.0E+18
 !! PRO 73
assign (resid 72 and name c ) (resid 73 and name n )
      (resid 73 and name ca) (resid 73 and name c )
      (resid 74 and name n )                61.32  30.34
 !! GLN 74
assign (resid 73 and name c ) (resid 74 and name n )
      (resid 74 and name ca) (resid 74 and name c )
      (resid 75 and name n )                52.15  27.80
 !! ASP 76
assign (resid 75 and name c ) (resid 76 and name n )
      (resid 76 and name ca) (resid 76 and name c )
      (resid 77 and name n )                54.29  41.07
 !! VAL 77
assign (resid 76 and name c ) (resid 77 and name n )
      (resid 77 and name ca) (resid 77 and name c )
      (resid 78 and name n )                58.82  1.0E+18
 !! PHE 78
assign (resid 77 and name c ) (resid 78 and name n )
      (resid 78 and name ca) (resid 78 and name c )
      (resid 79 and name n )                54.87  1.0E+18
 !! LEU 79
assign (resid 78 and name c ) (resid 79 and name n )
      (resid 79 and name ca) (resid 79 and name c )
      (resid 80 and name n )                51.98  40.79
 !! VAL 80
assign (resid 79 and name c ) (resid 80 and name n )
      (resid 80 and name ca) (resid 80 and name c )
      (resid 81 and name n )                59.79  29.27
 !! CYS 81
assign (resid 80 and name c ) (resid 81 and name n )
      (resid 81 and name ca) (resid 81 and name c )
      (resid 82 and name n )                56.18  28.74
 !! PHE 82
assign (resid 81 and name c ) (resid 82 and name n )
      (resid 82 and name ca) (resid 82 and name c )
      (resid 83 and name n )                54.83  1.0E+18
 !! SER 83
assign (resid 82 and name c ) (resid 83 and name n )
      (resid 83 and name ca) (resid 83 and name c )
      (resid 84 and name n )                53.07  61.60
 !! VAL 84
assign (resid 83 and name c ) (resid 84 and name n )
      (resid 84 and name ca) (resid 84 and name c )
      (resid 85 and name n )                62.10  28.59
 !! VAL 85
assign (resid 84 and name c ) (resid 85 and name n )
      (resid 85 and name ca) (resid 85 and name c )
      (resid 86 and name n )                58.35  28.45
 !! SER 86
assign (resid 85 and name c ) (resid 86 and name n )
      (resid 86 and name ca) (resid 86 and name c )
      (resid 87 and name n )                52.23  1.0E+18
 !! SER 88
assign (resid 87 and name c ) (resid 88 and name n )
      (resid 88 and name ca) (resid 88 and name c )
      (resid 89 and name n )                60.11  1.0E+18
 !! SER 89
assign (resid 88 and name c ) (resid 89 and name n )
      (resid 89 and name ca) (resid 89 and name c )
      (resid 90 and name n )                59.75  62.08
 !! PHE 90
assign (resid 89 and name c ) (resid 90 and name n )
      (resid 90 and name ca) (resid 90 and name c )
      (resid 91 and name n )                58.26  1.0E+18
 !! GLU 91
assign (resid 90 and name c ) (resid 91 and name n )
      (resid 91 and name ca) (resid 91 and name c )
      (resid 92 and name n )                57.17  27.17
 !! ASN 92
assign (resid 91 and name c ) (resid 92 and name n )
      (resid 92 and name ca) (resid 92 and name c )
      (resid 93 and name n )                52.98  36.19
 !! VAL 93
assign (resid 92 and name c ) (resid 93 and name n )
      (resid 93 and name ca) (resid 93 and name c )
      (resid 94 and name n )                66.71  29.99
 !! LYS 94
assign (resid 93 and name c ) (resid 94 and name n )
      (resid 94 and name ca) (resid 94 and name c )
      (resid 95 and name n )                56.55  31.80
 !! GLU 95
assign (resid 94 and name c ) (resid 95 and name n )
      (resid 95 and name ca) (resid 95 and name c )
      (resid 96 and name n )                55.91  28.79
```

```
!! LYS 96
assign (resid 95 and name c ) (resid 96 and name n )
      (resid 96 and name ca) (resid 96 and name c )
      (resid 97 and name n )                 55.49  29.62
 !! TRP 97
assign (resid 96 and name c ) (resid 97 and name n )
      (resid 97 and name ca) (resid 97 and name c )
      (resid 98 and name n )                 59.25  1.0E+18
 !! VAL 98
assign (resid 97 and name c ) (resid 98 and name n )
      (resid 98 and name ca) (resid 98 and name c )
      (resid 99 and name n )                 66.94  1.0E+18
 !! PRO 99
assign (resid 98 and name c ) (resid 99 and name n )
      (resid 99 and name ca) (resid 99 and name c )
      (resid 100 and name n )                64.99  29.11
 !! GLU 100
assign (resid 99 and name c ) (resid 100 and name n )
      (resid 100 and name ca) (resid 100 and name c )
      (resid 101 and name n )                59.06  1.0E+18
 !! ILE 101
assign (resid 100 and name c ) (resid 101 and name n )
      (resid 101 and name ca) (resid 101 and name c )
      (resid 102 and name n )                64.36  34.34
 !! THR 102
assign (resid 101 and name c ) (resid 102 and name n )
      (resid 102 and name ca) (resid 102 and name c )
      (resid 103 and name n )                63.27  67.92
 !! HIS 103
assign (resid 102 and name c ) (resid 103 and name n )
      (resid 103 and name ca) (resid 103 and name c )
      (resid 104 and name n )                56.60  26.57
 !! HIS 104
assign (resid 103 and name c ) (resid 104 and name n )
      (resid 104 and name ca) (resid 104 and name c )
      (resid 105 and name n )                57.60  31.23
 !! CYS 105
assign (resid 104 and name c ) (resid 105 and name n )
      (resid 105 and name ca) (resid 105 and name c )
      (resid 106 and name n )                50.77  1.0E+18
 !! PRO 106
assign (resid 105 and name c ) (resid 106 and name n )
      (resid 106 and name ca) (resid 106 and name c )
      (resid 107 and name n )                63.45  30.87
 !! LYS 107
assign (resid 106 and name c ) (resid 107 and name n )
      (resid 107 and name ca) (resid 107 and name c )
      (resid 108 and name n )                53.36  30.67
 !! THR 108
assign (resid 107 and name c ) (resid 108 and name n )
      (resid 108 and name ca) (resid 108 and name c )
      (resid 109 and name n )                59.74  1.0E+18
 !! PRO 109
assign (resid 108 and name c ) (resid 109 and name n )
      (resid 109 and name ca) (resid 109 and name c )
      (resid 110 and name n )                60.81  32.00
 !! PHE 110
assign (resid 109 and name c ) (resid 110 and name n )
      (resid 110 and name ca) (resid 110 and name c )
      (resid 111 and name n )                52.90  42.03
 !! LEU 111
assign (resid 110 and name c ) (resid 111 and name n )
      (resid 111 and name ca) (resid 111 and name c )
      (resid 112 and name n )                52.44  42.25
 !! LEU 112
assign (resid 111 and name c ) (resid 112 and name n )
      (resid 112 and name ca) (resid 112 and name c )
      (resid 113 and name n )                52.10  42.39
 !! VAL 113
assign (resid 112 and name c ) (resid 113 and name n )
      (resid 113 and name ca) (resid 113 and name c )
      (resid 114 and name n )                58.02  32.22
 !! GLY 114
assign (resid 113 and name c ) (resid 114 and name n )
      (resid 114 and name ca) (resid 114 and name c )
      (resid 115 and name n )                47.57  0.0
 !! GLN 116
assign (resid 115 and name c ) (resid 116 and name n )
      (resid 116 and name ca) (resid 116 and name c )
      (resid 117 and name n )                54.55  24.45
 !! ILE 117
assign (resid 116 and name c ) (resid 117 and name n )
      (resid 117 and name ca) (resid 117 and name c )
      (resid 118 and name n )                63.05  34.74
 !! ASP 118
assign (resid 117 and name c ) (resid 118 and name n )
      (resid 118 and name ca) (resid 118 and name c )
      (resid 119 and name n )                53.81  39.78
 !! LEU 119
assign (resid 118 and name c ) (resid 119 and name n )
      (resid 119 and name ca) (resid 119 and name c )
      (resid 120 and name n )                53.67  41.40
 !! ARG 120
assign (resid 119 and name c ) (resid 120 and name n )
      (resid 120 and name ca) (resid 120 and name c )
      (resid 121 and name n )                58.48  28.49
 !! ASP 121
assign (resid 120 and name c ) (resid 121 and name n )
      (resid 121 and name ca) (resid 121 and name c )
      (resid 122 and name n )                51.15  39.70
 !! ASP 122
assign (resid 121 and name c ) (resid 122 and name n )
      (resid 122 and name ca) (resid 122 and name c )
      (resid 123 and name n )                49.90  41.54
 !! PRO 123
assign (resid 122 and name c ) (resid 123 and name n )
      (resid 123 and name ca) (resid 123 and name c )
      (resid 124 and name n )                64.02  30.48
 !! SER 124
assign (resid 123 and name c ) (resid 124 and name n )
      (resid 124 and name ca) (resid 124 and name c )
      (resid 125 and name n )                60.45  60.53
 !! THR 125
assign (resid 124 and name c ) (resid 125 and name n )
      (resid 125 and name ca) (resid 125 and name c )
      (resid 126 and name n )                66.02  1.0E+18
 !! ILE 126
assign (resid 125 and name c ) (resid 126 and name n )
      (resid 126 and name ca) (resid 126 and name c )
      (resid 127 and name n )                63.34  35.56
 !! GLU 127
assign (resid 126 and name c ) (resid 127 and name n )
      (resid 127 and name ca) (resid 127 and name c )
      (resid 128 and name n )                57.57  27.72
 !! LYS 128
assign (resid 127 and name c ) (resid 128 and name n )
      (resid 128 and name ca) (resid 128 and name c )
      (resid 129 and name n )                58.25  30.42
 !! LEU 129
assign (resid 128 and name c ) (resid 129 and name n )
      (resid 129 and name ca) (resid 129 and name c )
      (resid 130 and name n )                56.28  1.0E+18
 !! ALA 130
assign (resid 129 and name c ) (resid 130 and name n )
      (resid 130 and name ca) (resid 130 and name c )
      (resid 131 and name n )                53.67  15.95
 !! LYS 131
assign (resid 130 and name c ) (resid 131 and name n )
      (resid 131 and name ca) (resid 131 and name c )
      (resid 132 and name n )                57.46  30.49
 !! ASN 132
assign (resid 131 and name c ) (resid 132 and name n )
      (resid 132 and name ca) (resid 132 and name c )
      (resid 133 and name n )                50.46  37.63
 !! LYS 133
assign (resid 132 and name c ) (resid 133 and name n )
      (resid 133 and name ca) (resid 133 and name c )
      (resid 134 and name n )                55.57  26.82
 !! GLN 134
assign (resid 133 and name c ) (resid 134 and name n )
      (resid 134 and name ca) (resid 134 and name c )
      (resid 135 and name n )                52.19  31.74
 !! LYS 135
assign (resid 134 and name c ) (resid 135 and name n )
      (resid 135 and name ca) (resid 135 and name c )
      (resid 136 and name n )                52.39  31.08
 !! PRO 136
assign (resid 135 and name c ) (resid 136 and name n )
      (resid 136 and name ca) (resid 136 and name c )
      (resid 137 and name n )                60.70  30.53
 !! ILE 137
assign (resid 136 and name c ) (resid 137 and name n )
      (resid 137 and name ca) (resid 137 and name c )
      (resid 138 and name n )                58.32  35.45
 !! THR 138
assign (resid 137 and name c ) (resid 138 and name n )
      (resid 138 and name ca) (resid 138 and name c )
      (resid 139 and name n )                57.70  66.44
 !! PRO 139
assign (resid 138 and name c ) (resid 139 and name n )
      (resid 139 and name ca) (resid 139 and name c )
      (resid 140 and name n )                63.90  29.58
 !! GLU 140
assign (resid 139 and name c ) (resid 140 and name n )
      (resid 140 and name ca) (resid 140 and name c )
      (resid 141 and name n )                58.54  26.95
 !! THR 141
assign (resid 140 and name c ) (resid 141 and name n )
      (resid 141 and name ca) (resid 141 and name c )
      (resid 142 and name n )                  64.44  67.53
 !! ALA 142
assign (resid 141 and name c ) (resid 142 and name n )
      (resid 142 and name ca) (resid 142 and name c )
      (resid 143 and name n )                  53.71  15.48
 !! GLU 143
assign (resid 142 and name c ) (resid 143 and name n )
      (resid 143 and name ca) (resid 143 and name c )
      (resid 144 and name n )                  57.84  27.66
 !! LYS 144
assign (resid 143 and name c ) (resid 144 and name n )
      (resid 144 and name ca) (resid 144 and name c )
      (resid 145 and name n )                  57.82  30.42
 !! LEU 145
assign (resid 144 and name c ) (resid 145 and name n )
      (resid 145 and name ca) (resid 145 and name c )
      (resid 146 and name n )                  56.16  38.96
 !! ALA 146
assign (resid 145 and name c ) (resid 146 and name n )
      (resid 146 and name ca) (resid 146 and name c )
      (resid 147 and name n )                  53.84  16.45
 !! ARG 147
assign (resid 146 and name c ) (resid 147 and name n )
      (resid 147 and name ca) (resid 147 and name c )
      (resid 148 and name n )                  57.39  28.37
 !! ASP 148
assign (resid 147 and name c ) (resid 148 and name n )
      (resid 148 and name ca) (resid 148 and name c )
      (resid 149 and name n )                  55.81  38.66
 !! LEU 149
assign (resid 148 and name c ) (resid 149 and name n )
      (resid 149 and name ca) (resid 149 and name c )
      (resid 150 and name n )                  52.56  39.79
 !! LYS 150
assign (resid 149 and name c ) (resid 150 and name n )
      (resid 150 and name ca) (resid 150 and name c )
      (resid 151 and name n )                  55.80  1.0E+18
 !! ALA 151
assign (resid 150 and name c ) (resid 151 and name n )
      (resid 151 and name ca) (resid 151 and name c )
      (resid 152 and name n )                  49.57  18.52
 !! VAL 152
assign (resid 151 and name c ) (resid 152 and name n )
      (resid 152 and name ca) (resid 152 and name c )
      (resid 153 and name n )                  64.43  31.05
 !! LYS 153
assign (resid 152 and name c ) (resid 153 and name n )
      (resid 153 and name ca) (resid 153 and name c )
      (resid 154 and name n )                  53.31  30.19
 !! TYR 154
assign (resid 153 and name c ) (resid 154 and name n )
      (resid 154 and name ca) (resid 154 and name c )
      (resid 155 and name n )                  54.35  38.16
 !! VAL 155
assign (resid 154 and name c ) (resid 155 and name n )
      (resid 155 and name ca) (resid 155 and name c )
      (resid 156 and name n )                  56.78  33.23
 !! GLU 156
assign (resid 155 and name c ) (resid 156 and name n )
      (resid 156 and name ca) (resid 156 and name c )
      (resid 157 and name n )                  52.10  30.21
 !! CYS 157
assign (resid 156 and name c ) (resid 157 and name n )
      (resid 157 and name ca) (resid 157 and name c )
      (resid 158 and name n )                  54.09  1.0E+18
 !! SER 158
assign (resid 157 and name c ) (resid 158 and name n )
      (resid 158 and name ca) (resid 158 and name c )
      (resid 159 and name n )                  54.15  63.39
 !! ALA 159
assign (resid 158 and name c ) (resid 159 and name n )
      (resid 159 and name ca) (resid 159 and name c )
      (resid 160 and name n )                  52.93  17.47
 !! LEU 160
assign (resid 159 and name c ) (resid 160 and name n )
      (resid 160 and name ca) (resid 160 and name c )
      (resid 161 and name n )                  55.91  41.51
 !! THR 161
assign (resid 160 and name c ) (resid 161 and name n )
      (resid 161 and name ca) (resid 161 and name c )
      (resid 162 and name n )                  59.79  68.17
 !! GLN 162
assign (resid 161 and name c ) (resid 162 and name n )
      (resid 162 and name ca) (resid 162 and name c )
      (resid 163 and name n )                  57.49  24.05
 !! LYS 163
assign (resid 162 and name c ) (resid 163 and name n )
      (resid 163 and name ca) (resid 163 and name c )
      (resid 164 and name n )                  56.44  30.50
 !! GLY 164
assign (resid 163 and name c ) (resid 164 and name n )
      (resid 164 and name ca) (resid 164 and name c )
      (resid 165 and name n )                  44.64  0.0
 !! LEU 165
assign (resid 164 and name c ) (resid 165 and name n )
      (resid 165 and name ca) (resid 165 and name c )
      (resid 166 and name n )                  57.12  41.92
 !! LYS 166
assign (resid 165 and name c ) (resid 166 and name n )
      (resid 166 and name ca) (resid 166 and name c )
      (resid 167 and name n )                  58.55  30.43
 !! ASN 167
assign (resid 166 and name c ) (resid 167 and name n )
      (resid 167 and name ca) (resid 167 and name c )
      (resid 168 and name n )                  55.01  37.04
 !! VAL 168
assign (resid 167 and name c ) (resid 168 and name n )
      (resid 168 and name ca) (resid 168 and name c )
      (resid 169 and name n )                  65.40  29.50
 !! PHE 169
assign (resid 168 and name c ) (resid 169 and name n )
      (resid 169 and name ca) (resid 169 and name c )
      (resid 170 and name n )                  60.35  40.64
 !! ASP 170
assign (resid 169 and name c ) (resid 170 and name n )
      (resid 170 and name ca) (resid 170 and name c )
      (resid 171 and name n )                  55.84  38.16
 !! GLU 171
assign (resid 170 and name c ) (resid 171 and name n )
      (resid 171 and name ca) (resid 171 and name c )
      (resid 172 and name n )                  56.39  1.0E+18
 !! ALA 172
assign (resid 171 and name c ) (resid 172 and name n )
      (resid 172 and name ca) (resid 172 and name c )
      (resid 173 and name n )                  53.78  16.00
 !! ILE 173
assign (resid 172 and name c ) (resid 173 and name n )
      (resid 173 and name ca) (resid 173 and name c )
      (resid 174 and name n )                  64.18  35.97
 !! LEU 174
assign (resid 173 and name c ) (resid 174 and name n )
      (resid 174 and name ca) (resid 174 and name c )
      (resid 175 and name n )                  56.43  39.14
 !! ALA 175
assign (resid 174 and name c ) (resid 175 and name n )
      (resid 175 and name ca) (resid 175 and name c )
      (resid 176 and name n )                  52.36  17.16
 !! ALA 176
assign (resid 175 and name c ) (resid 176 and name n )
      (resid 176 and name ca) (resid 176 and name c )
      (resid 177 and name n )                  51.91  17.97
 !! LEU 177
assign (resid 176 and name c ) (resid 177 and name n )
      (resid 177 and name ca) (resid 177 and name c )
      (resid 178 and name n )                  53.47  40.95
 !! GLU 178
assign (resid 177 and name c ) (resid 178 and name n )
      (resid 178 and name ca) (resid 178 and name c )
      (resid 179 and name n )                  52.55  28.04
 !! PRO 179
assign (resid 178 and name c ) (resid 179 and name n )
      (resid 179 and name ca) (resid 179 and name c )
      (resid 180 and name n )                  59.98  29.35
 !! PRO 180
assign (resid 179 and name c ) (resid 180 and name n )
      (resid 180 and name ca) (resid 180 and name c )
      (resid 181 and name n )                  61.34  30.25
 !! GLU 181
assign (resid 180 and name c ) (resid 181 and name n )
      (resid 181 and name ca) (resid 181 and name c )
      (resid 182 and name n )                  52.64  28.31
 !! PRO 182
assign (resid 181 and name c ) (resid 182 and name n )
      (resid 182 and name ca) (resid 182 and name c )
      (resid 183 and name n )                  61.60  30.33
 !! LYS 183
assign (resid 182 and name c ) (resid 183 and name n )
      (resid 183 and name ca) (resid 183 and name c )
      (resid 184 and name n )                  54.61  31.18
 !! LYS 184
assign (resid 183 and name c ) (resid 184 and name n )
      (resid 184 and name ca) (resid 184 and name c )
      (resid 185 and name n )                  54.53  31.46
 !! SER 185
assign (resid 184 and name c ) (resid 185 and name n )
     (resid 185 and name ca) (resid 185 and name c )
     (resid 186 and name n )                  56.68  62.18
 !! ARG 186
assign (resid 185 and name c ) (resid 186 and name n )
     (resid 186 and name ca) (resid 186 and name c )
     (resid 187 and name n )                  54.49  29.08
 !! ARG 187
!!assign (resid 186 and name c ) (resid 187 and name n )
!!     (resid 187 and name ca) (resid 187 and name c )
!!                                        55.98  29.87
```