05-Molecular Modelling and Design for Proteins and Drugs

After scanning the sequence against the library of 3D profile fragments, each residue in the given sequence normally has numerous matches with residues from different profile fragments with different scores. The score indicates the preference of this residue adopting the conformation of the matched residue in the profile fragment. We can make a scatter plot of the main chain conformation (ϕ and ψ angles) of all the matched residues from the profile fragments in the library. The number of residues which fall into each of the three regions (α-helix, β-strand or coil) are calculated, weighted by the scores matched by that fragment. The distribution of the three states for all the structures in the profile fragment library are also calculated. The preference of each residue in one of the three conformation states is calculated using a statistical inference algorithm. The secondary structure state of a residue is assigned to be the one with the highest preference. Tests using 40 structurally non-homologous structures indicate that the method has an overall prediction accuracy of 64%.

Comparative modelling can be envisaged as two steps. The first is to solve the inverse folding problem: to define all those sequences that can adopt a particular fold. Operationally this is more usefully posed as defining whether a new sequence belongs to any of the known folds.It involves projecting restraints from a three-dimensional structure on to a one dimensional sequence. For this step we have calculated amino acid substitution tables in terms of local structural environmental parameters, which can be used to generate sequence templates for secondary structures, structural motifs and tertiary folds. The second step is to use the sequence, together with the knowledge that the protein belongs to a family of known fold, to construct a model. This form of protein modelling and prediction involves placing constrains from a known fold on a related protein sequence. The two steps require similar knowledge of the structures of protein families, and this knowledge can be expressed as rules that relate both local and global three-dimensional structure to patterns in the sequence of amino acids in a polypeptide chain. The method is comparative but exploits a broader knowledge-base of non-homologous protein structures.

MS-05.01.04 DESIGN STRATEGY FOR PROTEIN STABILITY: STABLE LOCAL CONFORMATION IN CONSISTENCY WITH THE GLOBAL CONFORMATION. By K. Ishikawa, S. Kimura, K. Morikawa, S. Kanaya and H. Nakamura, Protein Engineering Research Institute, 6-2-3 Furuedai, Suita, Osaka 565, Japan.

Various mechanisms of the protein stability have so far been studied, and several strategies to enhance the stability are now proposed. We have recently made lots of mutant proteins of ribonuclease HI from *Escherichia coli* (*E.coli* RNase HI, 17.6kDa), and studied the conformational stability and their crystal structures. As a result, several of the mutant proteins obtained the remarkable thermal stability, due to only very local amino acid replacements; For example, Lys95→Gly is considered to stabilize the local left-handed α-helical conformation by a Gly residue (S. Kimura *et al.*, *J. Biol. Chem.*, 1992, **267**, 22014-22017). His62→Pro may stabilize the short turn structure (K. Ishikawa *et al.*, *Protein Eng.*, 1993, **6**, 85-91). Val74→Leu, fills the cavity in the hydrophobic core (K. Ishikawa *et al.*, *Biochemistry*, 1993, in press). All three mechanisms are localized, and the characteristic features of individual amino acids contribute the increase of the thermal stability. Analyses of the crystal structures of the wild-type and mutant proteins of *E.coli* RNase HI less than 1.8Å resolution reveal that the global conformations of all those mutant proteins deviate very little from that of the wild-type protein. It means that those local structural changes can be permitted and even suitable for the original global conformation. The additivity of the mutations was confirmed (S. Kimura *et al.*, *J. Biol. Chem.*, 1992, **267**, 21535-21542), and the structural analysis of the associated protein from *Thermus thermophilus* show that these local mechanisms are used in the thermophilic protein (K. Ishikawa *et al.*, *J. Mol. Biol.*, 1993, **230**, 529-542).

MS-05.01.05 RULE-BASED APPROACHES TO COMPARATIVE MODELLING by Z.Y. Zhu , M. S.Johnson, H. Wako, R. Sowdhamini, N. Srinivasan, K. Guruprasad, Z.Sun, B. Reddy, S. Rufino, Y. Edwards, T. Blundell*. Imperial Cancer Research Fund Unit of Structural Molecular Biology, Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK

PS-05.01.06 COMPARATIVE STRUCTURAL AND STEREOELECTRONICAL STUDY OF PINACIDIL, DIAZOXIDE AND CROMAKALIM, POTASSIUM CHANNEL OPENERS BELONGING TO THREE DIFFERENT CHEMICAL CLASSES. By L. Dupont*[a], B. Pirotte, P. de Tullio, B. Masereel, M. Schynts and J. Delarge[b]. [a]Cristallographie, Institut de Physique, [b]Chimie Pharmaceutique, Institut de Pharmacie, University of Liège, B-4000 Liège, Belgium.

Over the past few years the number of chemical agents with K^+ channel opening properties has greatly expanded. They are separated into distinct chemical classes typically exemplified by cromakalim (a benzopyran), pinacidil (a pyridylalkyl-cyanoguanidine), diazoxide (a benzothiadiazine), nicorandil (a pyridinic nitro compound), minoxidil sulfate (a pyrimidine derivative) and RP 49356 (a pyridinic thioformamide). For the three best studied K^+ channel openers, the rank order of potency for vascular smooth muscle relaxation was found to be cromakalim > pinacidil > diazoxide whereas for their activity on insulin secreting cells, the order was diazoxide > pinacidil > cromakalim (Newgreen, D.T., Bray, K.M., Mellarg, A.D., Weston, A.H., Duty, S., Brown, B.S., Kay, P.B., Edwards, G., Longmore, J. and Southerton, J.S., Br. J. Pharmacol., 1990, 100, 605-615; Lebrun, P., Antoine, M.H., Devreux, V., Hermann, M., Herschuely, A., J. Pharmacol. Exp. Theor., 1990, 255, 948-954). Of particular interest is the intermediate position of pinacidil between diazoxide and cromakalim. Moreover, pinacidil could be regarded as a fairly good structural analog of diazoxide.
The present work tries to evaluate the level of structural analogy between these three classes of compounds by using crystallographic and infographic data.
A systematic search was performed with SYBYL (Tripos Associates Inc., St Louis, Missouri, USA) starting from the X-ray conformation of pinacidil optimized by the Tripos force field maximin2 energy minimizer. The analysis of the search process exhibits four interesting low energy conformations. The four selected geometries have been optimized using the semiempirical method AM1 (MOPAC 5.0) and have been compared in terms of total energy calculation : the lowest energy conformation is actually the one found in crystal. A comparative study was undertaken on conformations and stereoelectronical properties of pinacidil, diazoxide and cromakalim to highlight the similarities which could be related to

 Copyright © 1993 International Union of Crystallography

their common K^+ channel opening activity. Like pinacidil, the geometries of the latter molecules were also optimized by Tripos force field starting from their respective X-ray conformations. The isopotential maps of the optimized structure of diazoxide, cromakalim and the four low energy conformers of pinacidil were compared. The isopotential map of diazoxide shows a good analogy with that of two calculated conformations of pinacidil (both different from that in crystal). Recent studies (Schwanstecker, M., Brandt, C., Behrends, S., Schaupp, U. and Panten, U., Br. J. Pharmacol., 1992, 106, 295-301) have suggested that diazoxide and pinacidil could exert their activity on the pancreatic K_{ATP} channel after interaction with a common binding site. We tentatively conclude that the moderate activity of pinacidil on the insulin release from pancreatic B-cells could be explained by adoption for pinacidil of a low energy conformation which can be reasonably fitted on diazoxide both from the point of view of atom positions and of stereoelectronic properties.

PS-05.01.07 MODELLING OF SOME RETROVIRAL ASPARTYL PROTEINASES. STUDY OF THE SHORTER SEQUENCE REQUIRED FOR THE *IN VITRO* ENZYMATIC ACTIVITY. By S. Geoffre*, R. Léonard, S. Llido, P. Picard and G. Précigoux, Lab. de Cristallographie, University Bordeaux I, 33405 - Talence, France.

The Bovine Leukaemia Virus (BLV) and Human T-cell Leukaemia Virus (HTLV) aspartyl proteinases are reported as putative proteins made of 126 and 125 amino acids respectively ("long sequences"). Since all known aspartic proteinases contain a conserved active site and core structure, it is reasonable to attempt to use them to model the unknown structure of another aspartic proteinase. The crystal structures of Rous Sarcoma Virus (RSV PR) and Human Immunodeficiency Virus (HIV-1 PR) were used to align the sequences of BLV and HIV-1 PR and to construct models. These models show that BLV and HTLV-I proteinases made of only 116 and 115 amino acids respectively ("short sequences") display three dimensional structures similar to that observed for other retroviral proteinases. The ten amino acids of the carboxyl extremities of the BLV and HTLV-I "long sequences" do not have any equivalent residue in the alignment of the active RSV and HIV-1 PR and would not act in the catalysis process.

The real value of our models is underlined by the *in vitro* activity and inhibition of the synthetic BLV proteinase made of only 116 amino acids.

PS-05.01.08 DIRECTIONAL PREFERENCES OF BINDING OF FUNCTIONAL GROUPS. J.P. Glusker, C.W. Bock*, L. Shimoni, A. Kaufman, A.B. Carrell, H.L. Carrell, The Institute for Cancer Research 7701 Burholme Avenue, Philadelphia, PA 19111, USA, *Philadelphia College of Textiles and Science, Philadelphia, PA 19144, USA.

The directional preferences for binding of functional groups to different molecules are studied by use of the Cambridge Structural Database (CSD). Data on intermolecular interactions to a selected functional group and large numbers of small-molecule crystal structures are analysed. These are analyzed in a statistical manner by use of contoured scatterplots in order to determine the variability in such binding. The results can then be used for macromolecules, studied to lower resolution, in order to suggest modes of ligand binding. Two types of interactions will be described here.

Hydrogen bonding and metal binding to nitrogen-containing heterocycles has been studied in this way to give ranges within which such binding deviates from the plane of the ring system and from the line disecting the C-N-C angle. The results are compared with X-ray structural data at lower resolution for some acridine-oligonucleotide complexes and the surroundings of histidine rings in some protein crystal structures .

Additionally, the stereochemistry of binding of functional groups to metal cations, such as divalent magnesium, manganese and cobalt, will be described, and the relevance of such binding to enzyme mechanisms will be discussed.

Supported by grant GM 44360 from National Institutes of Health.

PS-05.01.09 MOLECULAR SIMULATION OF HUMAN INSULIN-LIKE GROWTH FACTOR I B. Xiao, C. Verma and G. G. Dodson
Department of Chemistry, University of York, York Y01 5DD, UK

Molecular modelling on human insulin-like growth factor I (IGF) has been carried out based on the X-ray structures of mini proinsulins and the IGF-I NMR structures using the computer system - Protein Workbench. Searches for equivalents to the C domain linking the A and B chains and to the D domain as an extension of the A chain C-terminal were made in the protein data bank. Molecular dynamics and minimization were carried out (using CHARMM) on a solvated system to determine the final model. The model of IGF's three-dimensional structure was compared with the X-ray structure of the mini proinsulin. It has been found that the A- and B- domains of the three dimensional model of IGF-I are consistent with molecule 1 of the mini proinsulin but not with the C-loop where the molecule of the proinsulin has only a short connection. The high mobility of D-domain was observed - not surprisingly. However, the geometry of the model is fairly good enough and the result was used to interpret the binding sites of IGF binding proteins.