PS03.04.16 LOCAL INDICATORS OF MODEL QUALITY
Zhou G., Blanc, E. & Chapman, M. S.; Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306-3015, USA.

Reciprocal space R-factors are a measure of overall model quality that are not sensitive to local detail. Jones *et al.* [*Acta Cryst.*, (1991), **A47**, 110-119] suggested the use of real space R-factors, calculated from a local comparison of observed and calculated electron density near individual residues, to detect severe errors such as frame shift of sequence assignment. We use random shifts while keeping good geometry to test the sensitivity of local R-factors to see whether more subtle errors can be detected. Previous random shift tests did not keep good geometry. Unfortunately, this changes the overlap of the electron density of covalently linked atoms and has substantial impact on the real space R-factors. This is unlikely to occur in actual protein structure determinations with the imposition of geometrical restrains. Detection of error is therefore more difficult than suggested by these previous tests. In the current work, tests with random shift models, in which good geometry is maintained, have shown that real space R-factors are correlated to introduced error, but weakly. The R-factors of the most incorrect regions rise above noise when averaged over several residues. We will present the results of systematic tests with maps and models of various qualities at various resolutions.

Computing V
Macromolecular Homology Modelling, Structural Families & Docking

MS03.05.01 COMPARISON OF SEQUENCES AND STRUCTURES.
W. Taylor, National Inst. For Medical Res., London NW7 1AA,UK

The construction of a model for a protein of unknown structure based on a similar sequence of known structure requires accurate alignment of the sequences. When the sequences are highly divergent it becomes necessary to use constraints derived from the known structure.

Structural constraints can be imposed at two levels: the simpler occurs locally in the sequence and can take the form of a bias to match known secondary structure with predicted secondary structure of like type, or to match regions of sequence variation with regions known to be exposed to solvent in the structure. These constraints can all be imposed during the alignment of the two sequences. The more complex constraints involve interactions that are non-local in the sequence - such as the preference to have cysteins sufficiently close to form a disulphide bridge, or hydrophobic residues packed in the core. Following the work of Eisenberg, simple constraints will be referred to as "1D/3D comparison" while the establishment of an alignment using true spatial interaction will be referred to as 'threading' (Jones, Taylor and Thornton, 1992, Nature, 358:86-89).

A dynamic programming algorithm can find the optimal alignment of A and B given a score relating all pairs of positions. To align a sequence with a structure, these scores would need to be a measure of how well, say, residue a_i fits in location b_j. This quantity cannot be obtained until the full environment around position b_j is determined. The problem is circular since the final alignment (or a good part of it) must be known in order to calculate it. The problem can be overcome by inverting the approach and asking not "how 'happy' is a_i on b_j?" but "if a_i is placed on b_j, then how 'happy' can it be made?". Consequently, the dynamic programming algorithm is applied to the matrix (ij**R**) in which each element ($ij\mathbf{R}_{mn}$) is a measure of the interaction of a^i (on b_j) with a_m on b_n. The best path through this matrix has an associated score, which provides a measure by which the validity of the original assumption (of placing a_i on b_j) can be assessed. These scores can be taken to form the higher-level matrix (**S**), from which a set of equivalences can be extracted by a dynamic programming calculation finding the highest scoring path compatible with a sequence alignment. Following the use of the same algorithm in protein structure comparison (Taylor and Orengo, 1989, JMB, 208:1-22), the individual scores ($ij\mathbf{R}_{mn}$) along the optimal path traced in each low-level matrix (ij**R**) were summed to produce the high-level matrix.

MS03.05.02 COMPARATIVE PROTEIN MODELING BY SATISFACTION OF SPATIAL RESTRAINTS.
Andrej Sali, Rockefeller University, 1230 York Avenue, New York, NY 10021, USA; e-mail:sali@rockvax.rockefeller.edu

Our approach to comparative protein modeling based on satisfaction of spatial restraints will be described and evaluated. In addition, the method will be illustrated by a combined modeling and site-directed mutagenesis study of heparin binding sites on mouse mast cell proteases. In the first stage of the method, the alignment between the sequence to be modelled and related template structures is obtained. In the second stage, restraints on various distances, angles, and dihedral angles in the sequence are derived from its alignment with the template structures. The restraints are expressed in the most general form as conditional probability density functions. Their form depends on tabulated correlations between sequence and structure as found in a database of 105 family alignments containing 416 homologous structures. And finally, the 3D model is obtained by minimizing violations of homology-derived and energy restraints, using conjugate gradients and molecular dynamics procedures. The derivation and satisfaction of spatial restraints have been implemented in the MODELLER program that is available by ftp from guitar.rockefeller.edu.

MS03.05.03 MODELING THE STRUCTURE OF NAD BOUND TO PERTUSSIS TOXIN.
Maxwell D. Cummings, Trevor N. Hart, Bart Hazes, and Randy J. Read. Departments of Biochemistry, and Medical Microbiology & Immunology, Universtiy of Alberta, Edmonton, AB, Canada, T6G 2H7. Ph:403-492-4696, Fax:403-492-7521,Email: max@clouseau.mmid.ualberta.ca.

We describe a novel application of the technique known as fragment-based ligand design. This method is usually applied to the prediction of novel ligands for target binding sites. Libraries of molecular fragments are docked to the binding site(s) of interest; subsequently, collections of the dockings are connected to create novel potential ligands. We have used several new flexible docking and superposition tools, as well as more conventional rigid-body methods, to examine NAD binding to the catalytic subunits of diphtheria and pertussis toxins. Docking simulations with the rigid fragments adenine and nicotinamide revealed that the low energy dockings clustered in three distinct sites on the two proteins. Two of the sites were common to both fragments. The structures of the dinucleotides adenylyl 3'-5' uridine 3' monophosphate, and, more recently, NAD, bound to diphtheria toxin reveal that one pair of adenine/nicotinamide docking clusters is consistent with the two related complex structures. We chose adenine/nicotinamide pairs of dockings from these clusters, and superimposed flexible models of NAD onto these pairs. A Monte Carlo-based flexible docking procedure and energy minimization were used to refine the modeled NAD-PT complexes. The modeled complexes account for sequence similarities between PT and DT, and are consistent with many results that suggest the catalytic importance of certain residues.