Computational conformational analysis of cyclohexaglycyl

David N J White and David H Kitson

Chemistry Department, The University, Glasgow G12 8QQ, UK

Conformational hyperspace searching procedures have been used in conjunction with molecular mechanics calculations to characterize a series of minimum energy cyclohexaglycyl conformations (MECs). The key MECs are identical to those found by Scheraga in an earlier study using different computational techniques. The gas phase global MEC is stabilized by numerous intramolecular hydrogen bonds between (solvent) inaccessible amide groups; in contrast to the solution and crystalline phase global MEC where all of the amide groups are accessible for solvent interactions and/or intermolecular hydrogen bonding. The conformational hyperspace searching procedures are also extremely useful in protein model building studies.

Keywords: computational conformational analysis, molecular mechanics, cyclohexaglycyl, conformational hyperspace

received 29 July 1985, revised 17 October 1985, accepted 21 October 1985

The conformational hyperspace of all but the simplest of molecules encompasses multiple local energy minima (e.g. the two *gauche* and one *anti* forms of n-butane). Molecular mechanics (MM) calculations will converge upon the MEC closest in conformational hyperspace to the starting point of the calculations; and will *not* simulate the climbing, or tunneling through, of energy barriers to reach the global MEC (*anti*-n-butane in this example) without external direction by the user, or the assistance of additional computational algorithms. Molecular dynamics and Monte Carlo calculations are capable of achieving what MM cannot, but they in turn have a different set of problems which vitiates their use as a simple and cheap alternative to MM calculations for exploring conformational space.

In order to ensure that the global MEC (GMEC), usually the conformation of interest, is reached by MM calculations multiple starting points are usually chosen. Random combinations of bond lengths, bond angles and torsion angles mapped onto a fixed molecular connectivity could be used as starting points, but this approach is unnecessarily thorough as only the torsion angles vary to a significant extent between MECS (it requires significantly more energy to alter the 'hard' variables like bond lengths and angles, in contrast to the 'soft' and therefore easily deformed torsion angles). The task of locating the conformations of interest can therefore be reduced to one of searching the torsion angle subspace of the molecule.

The location of low energy areas of torsion angle hyperspace is frequently achieved by means of the grid search technique. This is the simplest form of direct search optimization and the only one which is non-iterative. It involves the evaluation of a function (the steric energy) at a series of points chosen to blanket all areas of the conformational hyperspace of interest, followed by an examination of the function values to reveal the approximate positions of the MECs. Typically the steric energy would be evaluated for all combinations of 30° steps in each torsion angle, and the approximate MECs further characterized by MM calculations. Grid search is a viable proposal for problems involving a dimensionality of five or less, but computer time constraints render it impracticable for larger problems. Clearly an alternative procedure has to be found which is just as exhaustive as grid search, but much less expensive in terms of computer time, to deal with the problem of cyclohexaglycyl.

METHODS AND RESULTS

The computer program Glomin[1] was used to generate a family of low energy MECs for cyclohexaglycyl. Glomin is based on the generalization and refinement of a procedure developed by de Santis and Liquori for the theoretical conformational analysis of Gramicidin S[2]. Both techniques are described in full in the original papers so only a brief synopsis of the Glomin algorithm is given here. The steps involved in Glomin are shown in Figure 1.

The starting coordinates were obtained by building a model of the linear hexaglycyl main chain using the Mol[3] molecular graphics system, and the bond lengths and angles were chosen to be the equilibrium values incorporated in the peptide force-field[4] used in this study. The ω torsion angles were set to 180° and the φ, ψ values were arbitrary.

The generators are pairs of φ and ψ torsion angles which correspond to MECs derived from the Ramachandran map of the *N*-acetyl *N'*-methyl amide of each amino acid comprising the oligopeptide under consideration. In this instance the calculations are simplified by the fact that the residues are all Gly, but complicated by the fact that the generators in the Gly Ramachandran map are diffuse and ill defined. In order to sample all of the low-energy space in the *N*-acetyl *N'*-methyl glycyl amide map adequately the generators chosen were

 0263-7855/86/020112-05 $03.00

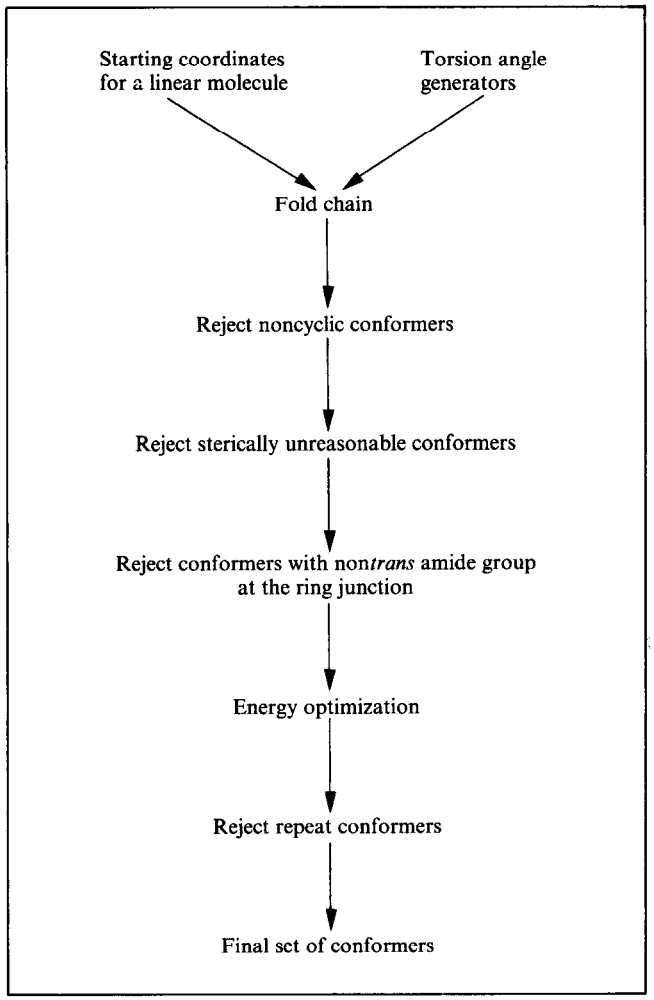

Figure 1. Schematic of the conformer generating program Glomin

(φ,ψ) = 1(90,60), 2(−120,60), 3(120,180), 4(−120,180), 5(120,−60), and 6(−90,−60).*

The hexaglycyl chain was folded into conformations corresponding to all possible (7776) combinations of the generators. Five pairs of φ,ψ torsion angles are permuted; the remaining φ,ψ pair and ω_6 remain undefined until the chain is folded and the chain ends joined after the creation of a successful cyclic conformation. It is obvious that of the large number of conformations calculated, only a small proportion will be even approximately cyclic. Advantage may be taken of this fact to eliminate noncyclic conformations at an early stage in the calculations. It is obvious that if any three consecutive φ,ω pairs of torsion angles are described by the generator sequence 333 (corresponding to a fairly extended main chain conformation) then no matter what values the other generators may assume, the ring will not close satisfactorily. Any number of these arbitrary length 'suicide sequences' (e.g. 333) may be input to Glomin and used to greatly reduce the computer time expended on the initial conformation generation step.

In order to cover conformational space in an area around the generator points, the calculated conformations were subjected to a torsion angle space optimization algorithm, with the chain end to chain end distance as the sole component of the objective function. Each torsion angle was allowed a latitude of $\pm 20^0$, and any conformation which could not achieve the ring closing distance range of 1.37 to 1.57 Å was eliminated from further consideration.

The surviving calculated conformations were then submitted to a battery of tests designed to eliminate unsatisfactory structures in a computationally inexpensive manner. These tests included: rejection of conformations which contained nonbonded distances less than 1.9 Å; rejection of conformations where ω_6 corresponds to a *cis*-amide bond; rejection of conformations where φ_6, ψ_6 correspond to high energy areas in the Ramachandran map; etc.

At this point there were 152 conformations remaining and these were prepared for MM calculations by adding on the carbonyl oxygens and the hydrogen atoms.

Initially all of the cyclohexaglycyl model conformations were given 110 cycles of pure-diagonal (PD) Newton–Raphson energy minimization, using Mol (on a minicomputer) to clean up the rather poor molecular geometries in the ring junction region of the molecules. The coordinates of these conformations were then stored on magnetic tape and transferred to an ICL 2976 mainframe computer for further MM calculations. (The authors did not have an array processor at this time and the calculations would have taken an inordinately long time on the PDP-11/40 minicomputer which hosts the Mol molecular graphics system). The structures were each allowed sufficient iterations (betwen 50 and 500) of block diagonal (BD) Newton–Raphson energy minimization[5] to reduce the r.m.s. force on the atoms below 0.05 kcal mol^{-1} $Å^{-1}$. Finally, the structures were given 3–10 cycles of full matrix (FM) Newton–Raphson energy minimization[5] using a generalized inverse technique to invert the Hessian matrix[6]. The r.m.s. force was then less than 10^{-8}–10^{-10} kcal mol^{-1} $Å^{-1}$ over the atoms of each conformation.

During the minimizations two conformations behaved abnormally. Conformer 35464 (string of generator digits) converged normally under the PD and BD procedures, but with FM the r.m.s. force stuck at 10^{-3} kcal mol^{-1} $Å^{-1}$ and would not go lower, despite repeated extra iterations. This behaviour is symptomatic of the minimizer having bottomed out in a very broad and ill-defined energy minimum. Conformer 21113 exhibited rather different symptoms which have since been observed on several other occasions. Again the difficulties started at the FM stage; where instead of a smooth diminution of the r.m.s. force, as is usually the case, it changed from 5×10^{-2} to 1 kcal mol^{-1} $Å^{-1}$ after 1 iteration, and the energy changed from −5 kcal mol^{-1} to 50 kcal mol^{-1}. This is not an infrequent occurrence when the model fed to the FM-MM is not close enough to the minimum for proper convergence; in these instances the calculation becomes completely unstable and the molecule 'blows apart'. However, what was

*For amino acids other than Gly there are a small number of well defined minima on their Ramachandran maps. The choice of generators and torsion angle latitude for ring closure to ensure full sampling of conformational space is therefore straightforward in these instances. In the case of Gly the situation is less simple; the authors therefore chose generators evenly distributed in conformational space and a torsion angle latitude as large as practicable. These choices have been tested by similar calculations on c-Gly_4, which presents a sterner test than c-Gly_6, and also where there are a large number of experimentally characterized conformations available to check the calculations[1].

unusual in the present case is that the forces and energy dropped on the second and subsequent iterations converging at 10^{-10} kcal mol^{-1} Å^{-1} and -8.74 kcal mol^{-1} respectively.

The explanation for the above behaviour is as follows: the PD minimizer begins with a conformation corresponding to the top of a potential energy well, and hands over to the BD minimizer some way down the well. The BD minimizer moves down the well until it comes to an *almost* horizontal *shelf* where the gradient lies between 10^{-8} and 5×10^{-2} kcal mol^{-1} Å^{-1} (but closer to the upper limit), at which point the BD algorithm converges. The FM minimizer moves down the not-quite-horizontal gradient and falls off the edge, whereupon both the energy and forces rise sharply. However, the atoms had been moving in the right direction, and provided that the gradient past the edge of the shelf is not too steep, then the FM algorithm can recover and ultimately converge.

It was then observed that in several cases more than one initially different generated conformation had converged upon the same MEC, or energy maximum. This is important and not universally recognized — gradient minimizers locate points in energy hyperspace where the net force on all atoms is zero. This condition is satisfied at energy maxima as well as minima. Fortunately it is easy to distinguish between the two cases when using FM NR optimization. If the 3N × 3N mass weighted Hessian matrix, of second derivatives of the steric energy with respect to the cartesian coordinates, is diagonalized there will be six zero-eigenvalues (corresponding to molecular translation and rotation) and 3N − 6 nonzero positive eigenvalues at an energy minimum; while for an energy maximum one or more of the eigenvalues will be negative.

After the repeat conformations had been deleted from the list 75 energy minima and maxima remained. The 20 lowest energy conformations are shown in Table 1. The separation of conformers into 'good' MECs and energy maxima was then attempted. This process is complicated by a number of things, for example; a small depression in the top of a profile corresponding to an energy maximum can show up as a MEC; or the top of a small ripple at the bottom of a large, shallow minimum can appear to be an energy maximum. Because oligopeptides are flexible, artifacts of this kind are distressingly common. After eliminating as many conformations of this kind as possible, and rerunning some calculations in quadruple precision as a safeguard against roundup errors in the numerical procedures, a total of 40 well defined MECs and energy maxima were located. A lowest-energy subset of these maxima is highlighted in Table 1 by an asterisk next to the conformer number. The first eight conformers of this lowest energy subset are shown in Figure 2.

Table 1. The generator sequences, calculated potential energies (kcal mol^{-1}), characterization, and symmetry of the calculated cyclohexaglycyl conformations. Multiple generator sequences for a conformer number indicates that the initial conformers described by the generator sequences converged upon the same minimum or maximum. Conformations marked with a * are 'good' MECs. Orthogonal cartesian coordinates for all good MECs are available from the authors on request.

Conformer number	Generator sequence	Relative energy	Extrema	Symmetry
1*	11312 21113 51131 66465 66466	0.00000	Min	C_2
2	25245	0.66833	Min	
3*	32632 32642 45135 45145 45245 51431 51455 52454 63223	1.25391	Min	C_2
4	63262 65245	1.26051	Min	
5	46523 52641	1.37538	Min	
6*	12645 32645 45132	1.52513	Min	C_i
7	25626 45123 45223 52231 52632 55463	1.92592	Min	
8	51312 65131	1.92940	Min	
9	26632 52615 66632	1.95504	Min	
10*	52515 52645	1.97018	Min	
11*	11431 24264 25431 25455 26412 45242 66412	2.19741	Min	
12*	11225 12232 51225	2.28092	Min	
13	45122 65132	2.45536	Max	
14	32513	2.49929	Max	
15*	12231 25131 31223 46514 55464 65626	2.77517	Min	
16 *	11131	2.79994	Min	
17 *	13262 15464	2.87494	Min	
18 *	31132 32633 32641 45146 46645	2.94646	Min	
19 *	12632	2.95553	Min	
20 *	12641 13264 51365 51366	2.96962	Min	

DISCUSSION

The most noticeable thing about Table 1 is the large number of conformations with values less than 3 kcal mol^{-1} higher in enthalpy than the GMEC. This observation goes a long way towards explaining the fact that the crystal structure of cyclohexaglycyl hemihydrate contains eight molecules in the asymmetric unit; with four molecules in one conformation (I, $\approx C_i$), two in another (II, C_1), and two unique conformations (III, C_i; IV, C_i) — all of which must be approximately isoenergetic[1]. On comparing the crystal structure conformations with the calculated results the most striking observation is the nonappearance of the calculated GMEC in the crystal structure.

Fortunately the nonobservation of the gas-phase GMEC (remember that the structures resulting from MM calculations correspond to absolute zero gas phase conformations) in the crystal structure is fairly easily rationalized. Figure 2 shows quite clearly that conformation 1 has a globular structure stabilized by a number of weak intramolecular hydrogen bonds, and that the orientation of the amide groups is far from ideal for the facile formation of intermolecular hydrogen bonds with other molecules of cyclohexaglycyl or water. On the other hand, conformation 3 of Figure 2 (corresponding to I of cyclohexaglycyl) is a very open structure consisting of two β-bends joined back to back. The amide groups in this structure are ideally placed to form strong hydrogen bonds with other molecules and solvent; so it comes as no surprise that conformations similar to 3 are observed in cyclohexaglycyl and a number of other crystal structures. While conformation 1 is favoured for the isolated molecule; conformation 3 more than makes good its higher intramolecular steric energy with strongly stabilizing intermolecular interactions in the liquid and crystalline phases. In point of fact conformation 1 has never been observed (nor is it likely to be, given the

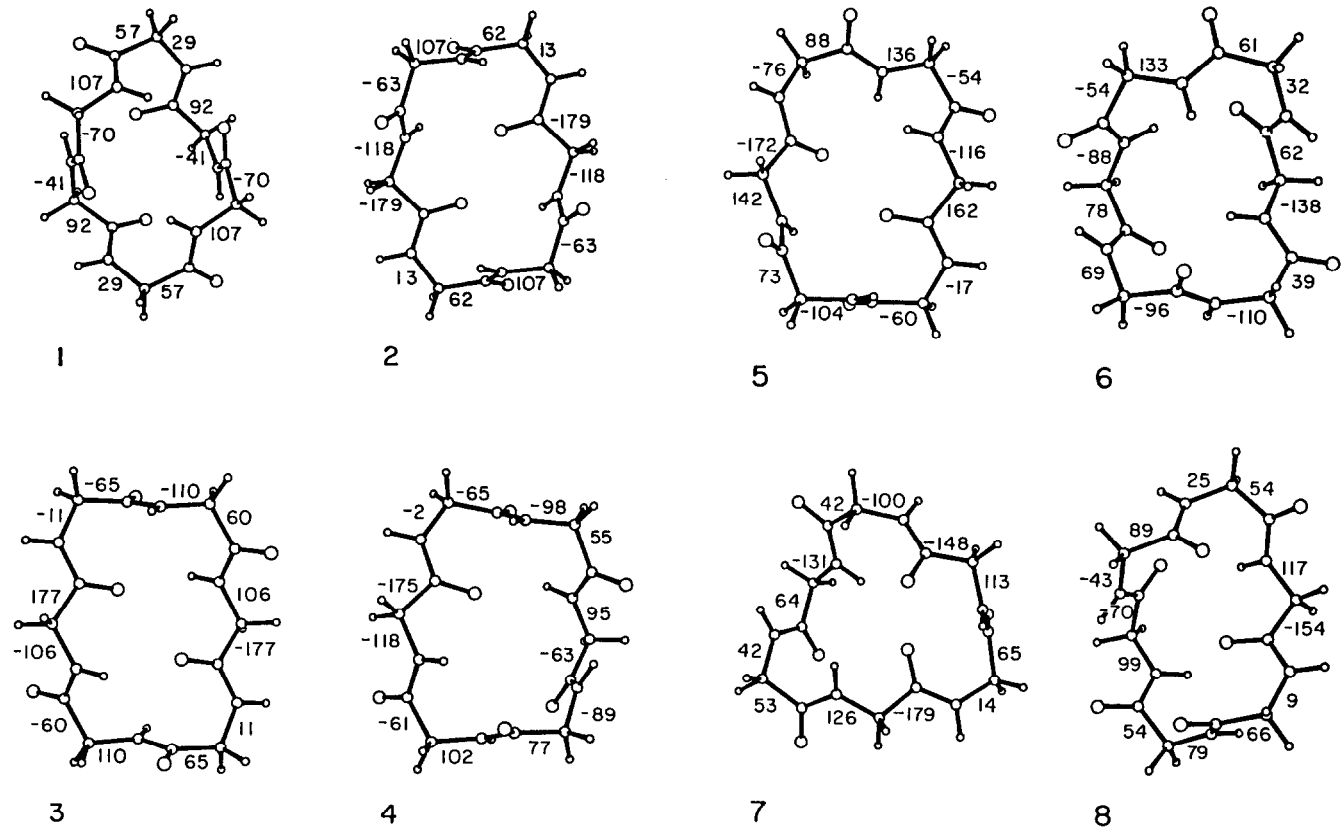

Figure 2. ORTEP diagrams of the eight lowest energy 'good' MECs calculated for cyclohexaglycyl. The φ and ω angles (°) are given. Calculated potential energies (kcal mol^{-1}) for the conformations are; conformation 1:0.00000; conformation 2:1.25391; conformation 3:1.52513; conformation 4:1.97018; conformation 5:2.19741; conformation 6:2.28092; conformation 7:2.77517; conformation 8:2.79994

low vapour pressure of cyclohexaglycyl and its insolubility in nonpolar solvents), and the only support for our calculations in this respect comes from the fact that Scheraga calculated the same gas phase GMEC using an entirely different conformer generating algorithm to ours[8].

Conformation IV in the crystal structure of cyclohexaglycyl is similar to conformation I and the average difference between corresponding torsion angles is only 19°. Crystal conformation II corresponds to calculated conformation 41 in the list of raw minima and maxima which has a calculated enthalpy of 3.61493 kcal mol^{-1} relative to the GMEC; while crystal conformation III corresponds to conformation 67 in the raw list with an enthalpy of 5.00293 kcal mol^{-1} above the GMEC. The average difference between corresponding calculated and observed torsion angles is in the region of 25° for conformations I–IV.

The magnitude of these differences is not entirely unexpected as the plethora of calculated MECs within 3 kcal mol^{-1} of each other and the multiple crystal structure conformations indicate that cyclohexaglycyl is probably a very flexible molecule. The occurrence of multiple calculated and observed low energy conformations provides no information about the activation energy necessary to interconvert conformations — the barriers could be high, but there could be multiple kinetically favoured routes to the various conformations. However, the PNMR spectrum of cyclohexaglycyl is consistent with a number of rapidly interconverting conformations[9], and there are three energy maxima, corresponding to transition states for conformational interconversions, less than 4.5 kcal mol^{-1} above the GMEC in our raw list of 75 calculated conformations. Given the fact that cyclohexaglycyl is very flexible and also that the crystal structure is a hydrate with multiple amide · · · water hydrogen bonds strongly influencing the molecular conformations, it is surprising that there is any correspondence at all between calculated and observed conformations.

Crystal structures of a number of other cyclic hexapeptides have been solved; in particular cyclo(L-Ala-L-Pro-D-Phe)$_2$[10], cyclo(Gly-L-Pro-D-Ala)$_2$[11], and cyclo(Gly-Pro-D-Phe)$_2$[12], all exhibit closely similar conformations with either C_2 or approximate C_2 symmetry. These in turn are broadly similar to the calculated conformation 2 (of Figure 2). The average difference between calculated and observed torsion angles is however greater than the corresponding difference for conformation 3. The three crystal structures mentioned above contain water, no solvate, and dimethyl sulphoxide respectively. These will exert differing forces on the cyclic hexapeptides. The side chains will also influence the main chain conformational angles, and the greater differences between calculated cyclohexaglycyl and observed mixed-residue cyclic hexapeptides is not too surprising.

There is however another factor which influences the agreement between calculated and observed conformations, and this is the quality of the force-field employed in the MM calculations — which is almost by definition inadequate. These observations suggest that one very effective means of optimizing polypeptide–protein force-fields would be calibration against cyclic hexapeptide crystal structures, where a good force-field

should be able to simulate the subtle interplay of intra- and interatomic forces on a flexible molecule. This exercise has recently been reported; currently popular peptide–protein force-fields were evaluated in terms of their ability to reproduce three reference cyclic hexapeptide crystal structures[13]. In the light of the above discussion this is the only way in which X-ray structures of flexible cyclic hexapeptides should be used to evaluate force-fields. If a single molecule has to be isolated for these purposes it must be rigid enough for its conformation not to be significantly affected by its environment; otherwise gas–solid phase conformational comparisons are meaningless. Crystal structures have been used in the past to develop, not test, polypeptide–protein force-fields but the molecules contained in the force-fields were either very small or very rigid[14] and so did not present a real challenge to the force field.

Despite the fact that a number of relatively high energy MECs were calculated that do not appear in Scheraga's list of cyclohexaglycyl MECs[8], and *vice-versa,* one common observation is that significant numbers of MECs are linked by amide flips (the rotation of an amide group about a vector joining the two adjacent α-carbon atoms). This suggests that generation of conformations by systematic permutations of amide flips might be a viable technique for global energy minimization of the cyclohexaglycyl model. In fact, the authors have successfully used such a procedure to generate MECs of cyclotri-β-alanyl[15] by permutations and combinations of $-CH_2-CH_2-$ twists and amide flips; where the dimethylenes were twisted either to the 'right' or 'left' (see Reference 15 for definition) and the amide groups were either 'up' or 'down' (the definition of 'up' or 'down' is arbitrary but one possibility could be: orient the mean plane of the starting conformation of the molecule to coincide with the *xy* plane of the coordinate system, and then 'up' corresponds to the amide C=O vector pointing roughly in the positive *z* direction and 'down' to a 180° rotation of the amide group from the 'up' position about the adjacent α-carbon atoms). Inspection of the cyclohexaglycyl conformations with the molecular graphics system suggests that the cyclotri-β-alanyl procedure might have to be modified to allow the planes of the amide groups to rotate by 90°, 180°, or 270° from either the 'up' or 'down' orientations, rather than by just 180°. This would lead to quite a small set of starting conformations (4096; less those related by symmetry to a previously generated conformation, those that were cyclic permutations of a previous conformer, and those that were easily recognized as sterically unreasonable) and therefore to less computer time. However, without actually performing the calculation there is no way of knowing for certain whether all of the important MECs would be found; but in a recent molecular dynamics study of cyclic enkephalin analogues[16] amide flips were actually observed, on a picosecond timescale, and found to be one of the 'softest' (and therefore most frequently occurring) transitions.

Shortly after this work was completed the authors began to build a model of human thrombin[17] from a knowledge of its amino-acid sequence[18], and the X-ray crystal structure of bovine α-chymotrypsin[19]. The amino-acid substitutions were dealt with in an automatic and straightforward way by the computer program developed for protein model building[20], as were the insertions which could be modelled from the crystal structures of elastase and trypsin using Greer's method[21]. The remaining one, two and three residue insertions/deletions were also handled automatically using a rule base and local energy minimizations. This left one 10 and two 5 residue insertions for which there was no precedent in the other homologous structures and no readily obtainable model in nonhomologous structures. In these instances models corresponding to the existing surface loops (where all insertions and deletions occurred) plus the inserted amino-acids are being investigated by Glomin. The normal ring-closing geometric constraints are replaced by constraints which ensure that the conformations generated by Glomin will fit into the protein structure.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge an SERC grant which enabled the purchase of molecular modelling hardware.

REFERENCES

1 **White, D N J and Morrow, C** *Comp. Chem.* Vol 3 (1979) p 33
2 **de Santis, P and Liquori, A M** *Biopolym.* Vol 10 (1971) p 699
3 **White, D N J and Kitson, D H** Unpublished work
4 **White, D N J and Guy, M H P** *J. Chem. Soc. Perkin II* (1975) p 43
5 **White, D N J** in *Specialist periodical reports — molecular structure by diffraction methods* Vol 6 (1978) p 38
6 **White, D N J** *Acta Cryst.* Vol A33 (1977) p 1010
7 **Karle, I L and Karle, J** *Acta Cryst.* Vol 16 (1963) p 969
8 **Go, N and Scheraga, H A** *Macromolecules,* Vol 11 (1978) p 552
9 **Tonelli, A E and Brewster, A I** *J. Am. Chem. Soc.* Vol 94 (1972) p 2851
10 **Brown, J N and Teller, R G** *J. Am. Chem. Soc.* Vol 98 (1976) p 7565
11 **Kostansek, E C, Lipscomb, W N and Thiessen, W E** *J. Am. Chem. Soc.* Vol 101 (1979) p 834
12 **Brown, J N and Yang, C H** *J. Am. Chem. Soc.* Vol 101 (1979) p 445
13 **Hall, D and Pavitt, N** *J. Comp. Chem.* Vol 5 (1984) p 441
14 **Hagler, A T, Huler, E and Lifson, S** *J. Am. Chem. Soc.* Vol 96 (1974) p 5319
15 **White, D N J, Morrow, C, Cox, P J, Drey, C N C and Lowbridge, J** *J. Chem. Soc. Perkin II* (1982) p 239
16 **Mammi, N J, Hassan, M and Goodman, M** *J. Am. Chem. Soc.* Vol 107 (1985) p 4008
17 **White, D N J** *Proc. Royal Soc. Series B* (1986) in press
18 **Friezner Degen, S J, MacGillivray, R T A and Davie, E W** *Biochem.* Vol 22 (1983) p 2087
19 **Birktoft, J J and Blow, D M** *J. Mol. Biol.* Vol 68 (1972) p 187
20 **White, D N J and Pearson, J E** *J. Mol. Graphics* to be submitted
21 **Greer, J** *J. Mol. Biol.* Vol 153 (1981) p 1027