# Some Aspects Concerning Conformation of Polypeptide Chains in Proteins

Alfred Gieren, Bernhard Dederer, and Friedrich Schanda

Max-Planck-Institut für Biochemie, Abteilung für Strukturforschung I, Am Klopferspitz, D-8033 Martinsried/München

Z. Naturforsch. **35 c,** 741–746 (1980); received April 17, 1980

NH · · · N Hydrogen Bridge, $\varphi/\psi$ Correlations, Nucleophilic Substitution at $C_\alpha$, $|\psi|/\tau$ Correlation

The structure of (S)-N,N′-di-tert-butyl-2-[N-(1-phenylethyl)benzamido] malonamide contains two fragments of a polypeptide chain. This compound therefore can be taken as a model substance for details of protein conformation. In the crystalline state one peptide chain of the model molecule incorporates a hydrogen bond between two adjacent nitrogen atoms in the backbone. The acceptor for the hydrogen is the $p_z$-orbital at the proton accepting nitrogen. The occurence of such hydrogen bonds in proteins might explain some correlations found between $\varphi$ and $\psi$ torsion angles. In addition a correlation between torsion angle $|\psi|$ and the bond angle $\tau$ at $C_\alpha$ in the backbone of polypeptide chains could be established. The model substance also contains a "frozen" back side attack of a C=O group on the tetrahedrally coordinated $C_\alpha$ in analogy to the $S_N2$ substitution reaction of the Walden inversion with a trigonal bipyramidal transition state.

Recently we determined by single crystal X-ray structure analysis the molecular structure of (S)-N,N′-di-*tert*-butyl-2-[N-(1-phenylethyl)benzamido]-malonamide (**1**) [1, 2], a byproduct of the 4CC-reaction [3, 4]. This compound contains two chemically identical fragments of a polypeptide chain – tripeptide fragments – which differ by torsion angles $\varphi$ and $\psi$.

The molecular structure of **1** is shown in Fig. 1. One polypeptide fragment starts from the C-terminal end with C1 ($C_\alpha$-atom) and follows the hatched chain to C15. The fragment C1 → C15 is analogous to a polypeptide fragment $C_\alpha$–NH–CO–$C_\alpha$–NH–CO–$C_\alpha$; the only difference is that N3 is substituted by a 1-phenylethyl residue – and not an H-atom – and C15 is $sp^2$-hybridized and not $sp^3$. The second polypeptide chain begins at C5 and follows the shaded chain to C3 and is, from C3 up to C15, identical with the first one. As can be seen from Fig. 1 the peptide groups are in the *trans* configuration usually found in proteins. Although $C_\alpha$3 is not chiral, for it is substituted by two equivalent substituents, the residue C1–C15 can be regarded to contain an L-amino acid and the residue C5–C15 an D one. The $\varphi/\psi$ torsion angle pair of the residue C1–C15 is + 70° and + 22° respectively and that of the residue C5–C15 is −64° and + 162° respectively. Because in this paper only stereochemical features of relative and not absolute conformation are discussed, for comparison with real protein structures D and L can be changed and also the signs of $\varphi/\psi$ by pairs. The $\varphi/\psi$ pairs found here occur in the so named $\beta$-turns [5–8], the torsion angle pair in the residue C1–C15 frequently and the pair in residue C5–C15 rarely.

The conformation of proteins is strongly influenced by hydrogen bonds, preferably by those between an N–H group as donor and a C=O group as acceptor. Therefore we first checked the molecular structure for the presence of intramolecular hydrogen bridges. Two hydrogen bonds can be found there. The N2–H2 group acting as a donor forms a hydrogen bridge to the C2=O1 carbonyl group acting as an acceptor. The H2 · · · O1 distance is 1.95 Å. This hydrogen bond is analogous to similar hydrogen bonds between neighbouring peptide chains in proteins.

Far more interesting is the second hydrogen bond within the peptide fragment C1 → C15 which also starts from an N–H group as donor but quite unexpectedly binds to an acylamide nitrogen acting as acceptor. The length of the H1 · · · N3 contact distance of 2.27 Å shows that this bridge must be a weak one, but the distance found here is significantly shorter than the sum of the van-der-Waals-radii of 2.7 Å [9]. The N1 · · · N3 distance with 2.74 Å is also shorter than the sum of van-der-Waals-radii (3.0 Å). For, due to the $\pi$-resonance in the peptide group, the basicity of the nitrogen, and therefore its

Reprint requests to Dr. A. Gieren.

0341-0382/80/0900-0741 $ 01.00/0

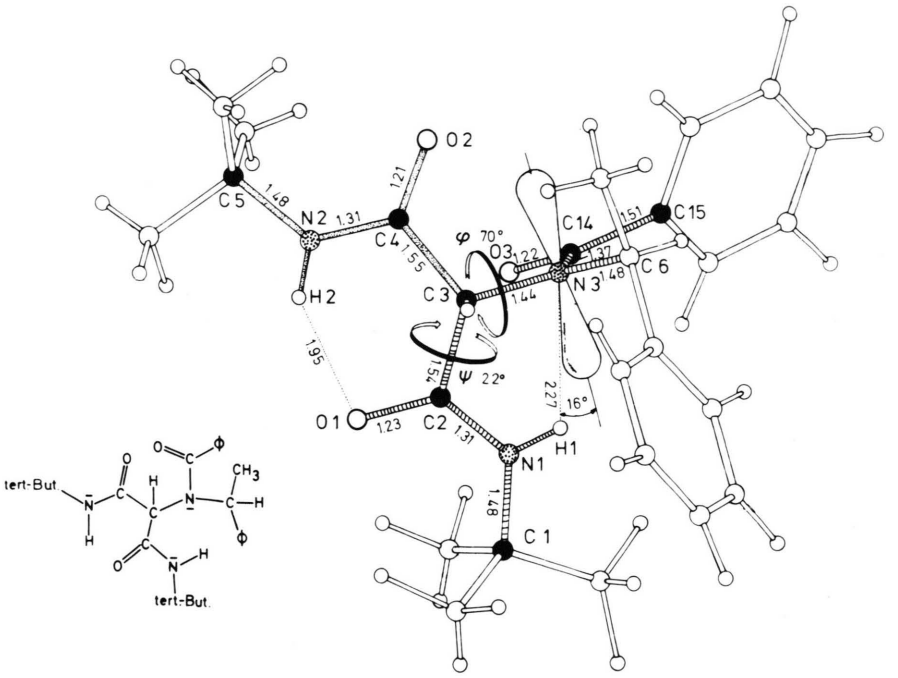
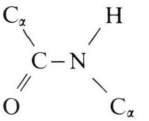



Fig. 1. Molecular structure of (S)-N,N′-di-*tert*-butyl-2-[N-(1-phenylethyl)benzamido]malonamide. Relevant bond distances are given.

acceptor property for protons, is decreased. However, our model compound shows that under certain steric conditions the acylamide nitrogen may yet function as a proton acceptor.

Fig. 1 illustrates that the proper acceptor for the partially positively charged hydrogen H1 is the $p_z$-orbital at the acceptor nitrogen N3. The angle between the H1 · · ·N3 vector and the axis of the $p_z$-orbital amounts to only 16°. The participation of the $p_z$-orbital in the hydrogen bond is substantiated by two bond lengths starting at the acceptor nitrogen N3. The peptide C–N-bond C14–N3 is 1.37 Å long and significantly elongated with relation to the other two analogous C–N bond lengths of C2–N1 and C4–N2 of 1.31 Å. However, the C3–N3 single bond of type $sp^3$–$sp^2$ with a length of 1.44 Å is short compared with the analogous C–N single bonds C6–N3, C1–N1 and C5–N2 which all have a length of 1.48 Å. The shortening of the C3–N3 bond length and the simultaneous elongation of the peptide bond C14–N3 may be interpreted as follows. C3 is substituted by two electron withdrawing C=O groups, which tend to make C3 more positive, thereby withdrawing electron density from the peptide C14–N3 bond into the C3–N3 single bond. An additional effect is the resonance formula (**a**) for the peptide bond with a lone pair of electrons at the nitrogen N3 obtaining an increased weight which includes an increase of the electron density within the $p_z$-orbital at N3.




Resonance formula (**a**).

If this $p_z$-orbital also participates in the hydrogen bridge, then the bonding strength of this hydrogen bond not only depends on the H1 · · · N3 hydrogen bond distance, but also on the orientation of the $p_z$-orbital. The hydrogen bond therefore should be strongest if first, the H · · · N distance has a minimum and second, the angle between the H · · ·N vector and the axis of the $p_z$-orbital is simultaneously at a minimum. This is the case if, as is approximately fulfilled in the model structure, the two adjacent peptide planes C1 → C3 and C3 → C15 are perpendicular to each other, and the $\psi$ torsion angle amounts to 0° and the $\varphi$ torsion angle to 90° or −90° respectively. $\psi$ is defined as the dihedral angle N3–C3–C2–N1 and $\varphi$ as the dihedral angle C2–C3–N3–C14. The torsion angles $\psi$ and $\varphi$ are,

as is known, those which determine the convolution of peptide chains. The angle $\psi$ determines the H1 ··· N3 contact distance which is a minimum for $\psi = 0°$, N3 being in the plane of the peptide group C1 → C3. When simultaneously $\varphi$ is 90°, the axis of the $p_z$ orbital lies in the same plane. If now the torsion angle $\psi$ is changed, then $\varphi$ also must change, if one wants to preserve the alignment of H1 ··· N3 with the $p_z$ orbital. An increase of the angle $\psi$ for instance rotates H1 upward from the plane of projection in Fig. 1, and the $p_z$ orbital can be made to follow this rotation by rotating in the $-\varphi$ direction. A rotation by $+\psi$ is thus followed by a simultaneous rotation by an angle of $-\varphi$. In the model compound we find an angle $\psi$ of 22° and an angle $\varphi$ of 70° in accordance with the assumption of the $p_z$ orbital being the proton acceptor. In this context it should be noted that a pair $+\varphi$, $+\psi$ is to be set equal to a pair $-\varphi$, $-\psi$ and also a pair $+\varphi$, $-\psi$ is to be set equal to $-\varphi$, $+\psi$, since such a change of sign is only a mirror image, but not a change of relative conformation which is exclusively evident for the orientation of the H ··· N vector relative to the axis of the $p_z$ orbital.

In order to test whether this hydrogen bond found in the model structure actually plays a part in protein structures, we have investigated the $\varphi/\psi$ correlation in four refined protein structures, the insect hemoglobin erythrocruorin [10], the endopeptidase trypsin [11], the basic pancreatic trypsin inhibitor (BPTI) [12] and the variable part of a Bence-Jones immunoglobulin [13]. These examples represent a wide variety of globular protein structures. Erythrocruorin for instance contains a great portion of α-helix structure, whereas in trypsin and BPTI the α-helix content is much smaller and finally the Bence-Jones immunoglobulin fragment chosen has no α-helix content but is built up of a large amount of pleated sheets. In Fig. 2 in the range of $\psi$ between −50° and +50°, the $\varphi/\psi$ pairs of these four proteins are plotted with $\varphi$ versus $\psi$. One sees clearly that the angles $\varphi$ decrease with increasing $\psi$. The regression line intercepts the ordinate at $\varphi \approx 90°$ at a value of the abscissa of $\psi = 0°$. Due to the course of this regression line one may assume with high probability that hydrogen bonds with N−H as donor and an adjacent acylamide nitrogen as acceptor, pointing from the C-terminal to N-terminal end of a peptide chain, may occur within a $\psi$ range of about 100° in protein structures.

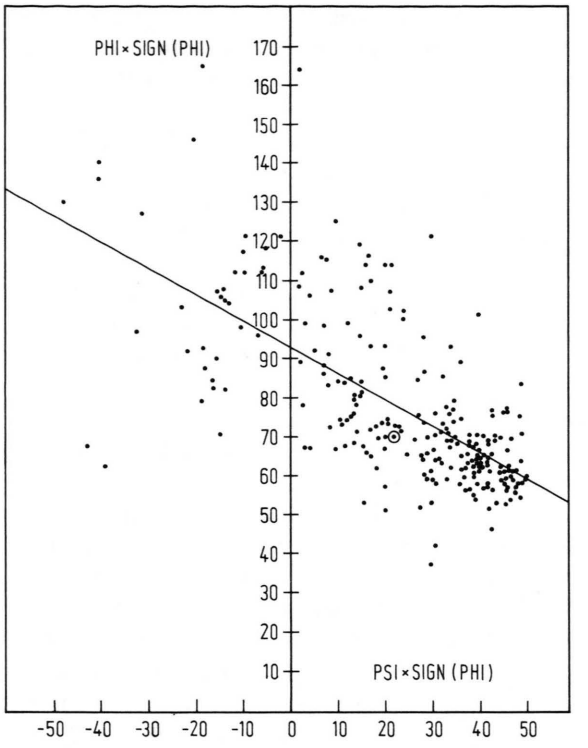



Fig. 2. Correlation between $\varphi$ and $\psi$ in a $\psi$-region of $-50° < \psi < 50°$ for the $\varphi/\psi$ pairs of erythrocruorin, trypsin, BPTI and the Bence-Jones immunoglobuline (variable part). As the crystal structure of the last protein contains two almost identical molecules, the $\varphi/\psi$ values are averaged. The equation of the regression line is: $\varphi \,\mathrm{sign}\,(\varphi) = -0.67\,\psi \,\mathrm{sign}\,(\varphi) + 92.8$. (The sign function is defined as follows: sign $(x) = -1, 0, +1$ for $x < 0$, $x = 0$, $x > 0$). The correlation coefficient amounts to −0.69. The $\varphi/\psi$ pair of the model-structure is marked with "○".

It should be mentioned that Pohl [14] found for 4 other protein structures and also Wu and Kabat [15] for 11 proteins a $\varphi/\psi$ distribution, which is analogous to that given in Fig. 2. Our Fig. 2 incorporates partially the regions A and C of a $\varphi/\psi$ plot given by Pohl [14]. It is also interesting to note that Pohl [14] found a relatively high density in an empirical protein energy map (EPEM) for regions $A_2$ and $C_2$ ($\varphi \sim \pm 90°$, $\psi \simeq 0°$). Pohl [14] interpreted this phenomenon by assuming "a partial positive charge $\delta^+$ at the amide proton interacts favourably with the delocalized π-electrons of the previous peptide bond". This interpretation is not far away from ours of an NH···N hydrogen bond. It is also interesting to note that Zimmermann *et al.* [16] in the course of conformational energy calculations using ECEPP (Empirical Conformational Energy Program for Pep-

tides) have improved the set of parameters by softening "the repulsive interaction between an amide nitrogen and the amide hydrogen in an other peptide group, in order to lower the energy in the "bridge region" (around $\varphi = -90°$, $\psi = 0°$). This lower energy is required since many residues in proteins have conformations in the "bridge region". The true reason for this may be the hydrogen bridge discussed here.

As mentioned before, the N–H · · · N hydrogen bond just discussed has the direction from the C-terminal to the N-terminal end of the polypeptide chain. The question arises whether hydrogen bonds of this type in the opposite direction are possible too, namely from the N-terminal to the C-terminal end of the peptide chain. In this case also the adjacent peptide planes have to be perpendicular to each other under ideal conditions, but now the $\varphi$ angle has to be 0° and the $\psi$ angle + or −90°. The arguments given before are also true for this hydrogen bridge, only $\varphi$ and $\psi$ have to be interchanged. In Fig. 3 we therefore have plotted $\psi$ against $\varphi$ in the $\varphi$-region $-50° < \varphi < 50°$ for the four proteins mentioned. As can be seen from this plot only a few $\varphi/\psi$ pairs occur in the protein structures which fulfill this condition. It is not surprising that so few residues occur here, for conformational energy calculations show that the region around $|\varphi| = 0$ is very unfavourable [16]. But the

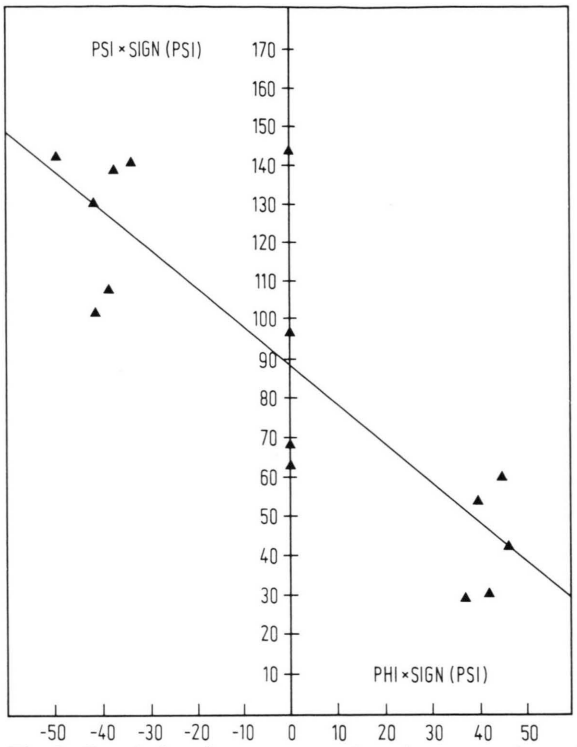



Fig. 3. Correlation between $\varphi$ and $\psi$ in a $\varphi$-region of $-50° < \varphi < 50°$ for $\varphi/\psi$ pairs of erythrocruorin, trypsin, BPTI and the Bence-Jones immunoglobuline. The $\varphi/\psi$ pairs of the two molecules in the crystal structure of the last are not averaged. The equation of the regression line is: $\psi$ sign $(\psi) = -0.99\,\varphi$ sign $(\psi) + 87.9$. The correlation coefficient amounts to −0.85.

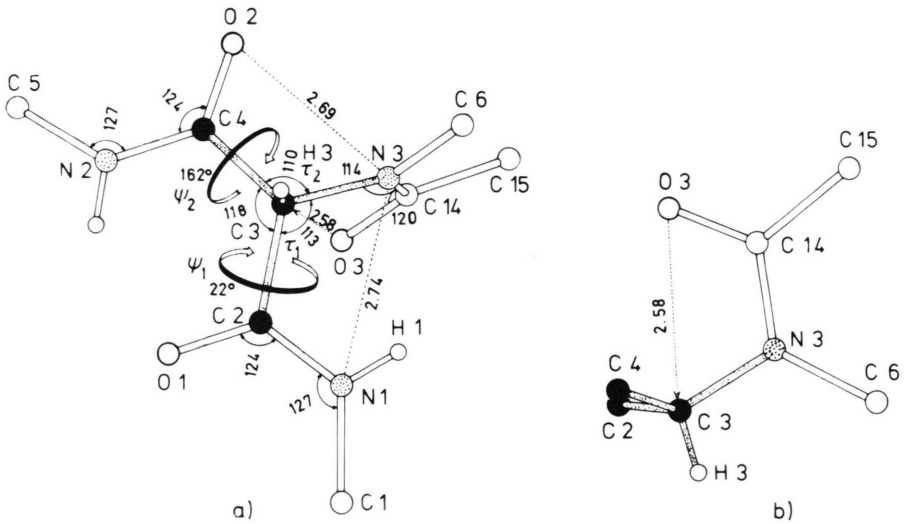



Fig. 4. a) Fragment of the molecular structure of (S)-N,N′-di-*tert*-butyl-2-[N-(1-phenylethyl)benzamido]malonamide. Relevant distances and angles are given. b) Distorted trigonal bipyramid at C3 viewed differently from a).

slope of the regression line and the intercept are as expected. The $\psi$ angle decreases with increasing $\varphi$ and the line goes nearly through the point $\varphi = 0°$ and $\psi = 90°$. Although this regression line is not as well defined as the analogous line of Fig. 2, the plot shows that if such unusual $\varphi/\psi$ pairs occur in protein structures, a hydrogen bond of type N–H···N in the direction from the N-terminal to the C-terminal end cannot be excluded.

Our model molecule shows another interesting peculiarity (Fig. 4). Within the same protein fragment containing the N–H···N hydrogen bridge we find a very short C3···O3 contact distance of 2.58 Å which is significantly shorter than the sum of the van-der-Waals-radii (3.1 Å). The bonding partners at C3 [C2, C4, N3, H3] as well as the carbonyl oxygen O3 form a distorted trigonal bipyramid with C3 as a center and H3 and O3 as apexes. Here we are dealing with a "frozen chemical reaction" [17, 18], namely the nucleophilic substitution of second order at a tetrahedral center with a trigonal bipyramidale transition state, the so-called Walden inversion. The oxygen acts as the entering and the hydrogen acts as the leaving group. This "frozen" back-side attack of the carbonyl group C14 = O3 is the reason for the $\varphi$ torsion angle with 70° deviating from the ideal value of 90° for the discussed hydrogen bridge. The short C3···O3 distance is obtained by decreasing the bond angles O3–C14–N3 and C3–N3–C14. These two bond angles with 120° and 114° are significantly smaller than the comparable angles at C2 and C4 (124°) and N1 and N2 (127°), respectively. Due to the bipyramidal "transition state" the tetrahedron at C3 is

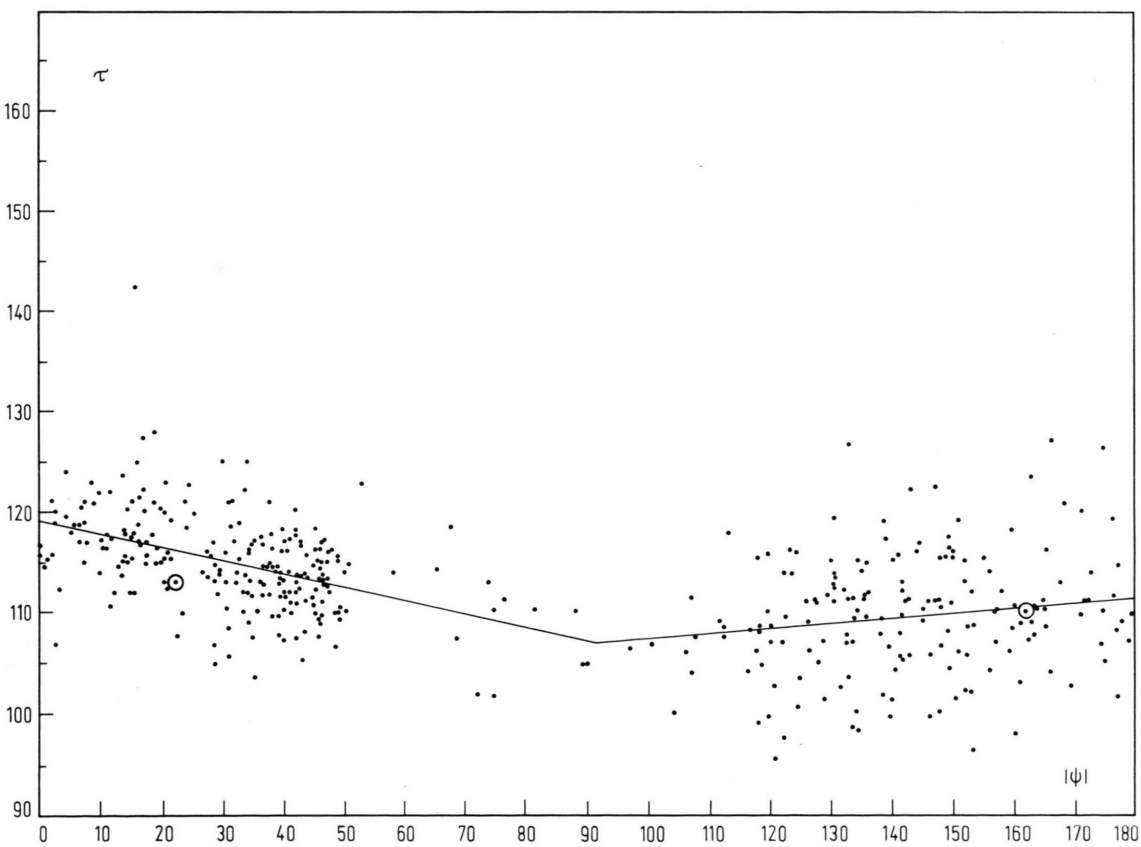



Fig. 5. Plot of $\tau/|\psi|$ pairs of erythrocruorin, trypsin and BPTI. The pairs of the model structure are marked with '○'. The equations of the regression lines are:

$$\tau = -0.13 \cdot |\psi| + 119.0 \quad (0° \leqq |\psi| < 90°) \quad \text{and}$$
$$\tau = \quad 0.05 \cdot |\psi| + 102.0 \; (90° \leqq |\psi| \leqq 180°).$$

The correlation coefficients amount to −0.47 and 0.16 respectively.

distorted in the direction of a trigonal pyramid. The angles C2–C3–C4 (118°) and C2–C3–N3 (113°) are greater than the tetrahedral angle and approach 120°. On the other hand the angles with H3, an apex of the trigonal bipyramid, are all smaller than 109.5° (101, 106, 108°), although the difference from 109.5° for a single value of this type is not significant within the limits of error. Up to now we have not been able to prove this "frozen" nucleophilic back-side attack to be present in protein structures, since some parameters important for this proof, for instance the angles at C and N of the peptide bonds, were not varied during refinement of the protein structures. On the other hand this "frozen" nucleophilic back-side attack in the model compound seems not to have strong relevance to protein structures, for the model contains two CO groups attached to the $C_\alpha$ atom C3, increasing the electrophilic character of this $C_\alpha$, what is not the case in real protein structures.

Looking for possible parameter correlations in protein structures in order to proof this phenomenon in proteins, we found a correlation between the bond angle at $C_\alpha$, the so called $\tau$ angle, and the amount of the torsion angle $\psi$, which is due to another steric effect. For the $\tau$ angle depends in the region $0° \leq |\psi| < 90°$, as is the case with the angle C2–C3–N2 (113°) ($|\psi_1| = 22°$), on the non-bonding interaction between two adjacent nitrogen atoms (N1, N3). The mutual repulsion of these two atoms is greatest at $\psi = 0°$ and smallest at $|\psi| = 90°$. Increasing the torsion angle $|\psi|$ beyond 90° substitutes the N···N interaction by an O···N interaction, as is the case between O2···N3 with a $|\psi_2|$ angle of 162° and a $\tau_2$ angle of 110°. This contact distance is minimized at $|\psi| = 180°$ and maximized at $|\psi| = 90°$. Therefore one should assume that the $\tau$ angle should be the smallest at $|\psi| = 90°$ and should become greater as a function of the deviation from $|\psi| = 90°$.

Fig. 5 shows the $\tau/|\psi|$ correlation of three proteins. Apparently, starting from $|\psi| = 90°$, the $\tau$ angles become greater in the increasing as well as in the decreasing direction of $|\psi|$. The distribution of $\tau/|\psi|$ pairs in first approximation can be described by two straight lines. The slope of the regression line seems to be greater in the region $0° \leq |\psi| < 90°$ than in the region $90° \leq |\psi| \leq 180°$, in agreement with stereochemical considerations. The values $\tau_1$ and $\tau_2$ found in the model structure fit very well the distribution found in proteins. In agreement with the correlation given here is that $\langle\tau\rangle$ usually is reported to be $\simeq 112°$, greater than the tetrahedral angle of 109.5°. It should be mentioned that the correlation plotted in Fig. 5 deteriorates if one takes into account, besides the pairs of the structure of erythrocruorin, trypsin and BPTI, also the $\tau/|\psi|$ pairs of the Bence-Jones immunoglobuline fragment. The explanation for this behaviour is that the resolution and refinement of the last protein is not at such a high level as is the case with the three other structures, for the $\tau$-angles are very sensitive to resolution of structure and level of refinement. From the correlation demonstrated in Fig. 5 one therefore can conclude that it is useful to vary the $\tau$-angles, at least in the last stages of refinement of protein structures.

*Acknowledgements*

The authors thank Prof. R. Huber and his coworkers for supplying the parameters of the cited protein structures.

[1] A. Gieren and B. Dederer, Tetrahedron Letters **1977,** 1503.
[2] A. Gieren and B. Dederer, Acta Cryst. **B 34,** 533; 2067 (1978).
[3] I. Ugi and G. Kaufhold, Liebigs Ann. Chem. **709,** 11 (1967).
[4] G. Gokel, G. Lüdke, and I. Ugi, Isonitrile Chemistry, (I. Ugi, ed.), Chapter 8, Academic Press, New York and London 1971.
[5] G. M. Venkatachalam, Biopolymers **6,** 1425 (1968).
[6] J. L. Crawford, W. N. Lipscomb, and C. G. Schellman, Proc. Nat. Acad. Sci. USA **70,** 538 (1973).
[7] P. N. Lewis, F. A. Momany, and H. A. Scheraga, Biochim. Biophys. Acta **303,** 211 (1973).
[8] P. Y. Chou and G. D. Fasman, J. Mol. Biol. **115,** 135 (1977).
[9] L. Pauling, The Nature of the Chemical Bond, p. 260, Cornell Univ. Press, Ithaca 1960.
[10] W. Steigemann and E. Weber, J. Mol. Biol. **127,** 309 (1979).
[11] W. Bode and P. Schwager, J. Mol. Biol. **98,** 693 (1975).
[12] J. Deisenhofer and W. Steigemann, Acta Cryst. **B 31,** 238 (1975).
[13] O. Epp, E. E. Lattman, M. Schiffer, R. Huber, and W. Palm, Biochemistry **14,** 4943 (1975).
[14] F. M. Pohl, Nature New Biol. **234,** 277 (1971).
[15] T. T. Wu and E. A. Kabat, J. Mol. Biol. **75,** 13 (1973).
[16] S. C. Zimmermann, M. S. Pottle, G. Némethy, and H. A. Scheraga, Macromol. **10,** 1 (1977).
[17] H. B. Bürgi, Inorg. Chem. **12,** 2321 (1973).
[18] H. B. Bürgi, J. D. Dunitz, and E. Shefter, Acta Cryst. **B 30,** 1517 (1974).