Theor Chim Acta (1995) 90: 41–50

Theoretica Chimica Acta

Transferability of the –COOH··· $^{-}$OOC– dyad Geometry from the gas phase to crystals and proteins

Tamás Nusser[1], László Turi[1], Gábor Náray-Szabó[1], Kálmán Simon[2]

[1] Department of Theoretical Chemistry, Eötvös University Budapest, P.O. Box 32, H-1518 Budapest 112, Hungary

[2] Department of Organic Chemical Technology, Technical University of Budapest H-1521 Budapest, Hungary

Received January 25, 1994/Accepted August 18, 1994

Summary. Simple models of the $-COO^{-} \cdots$ HOOC– moiety were studied by *ab initio* and semiempirical MNDO/PM3 methods combined with retrieval from the Cambridge Crystallographic Database and Protein Data Bank in order to check transferability of the non-bonded geometry. We found that the gas-phase O ··· O distance (252 pm) elongates in crystals and proteins to 256 ± 14 and 262 ± 27 pm, respectively. C–O ··· O–C dihedral angles are much more variable, however, energetically excluded values appear only exceptionally in crystals and less frequently in proteins. Most probably due to packing effects $-COO^{-} \cdots$ HOOC– conformations, preferred in the gas phase, are poorly populated in crystals

Key words: Carboxyl – Carboxylate dyad – Geometry – Transferability – Crystals – Proteins

Introduction

Quantum chemical study of enzymatic reaction paths needs a suitable method and an appropriate molecular model. The considerable development achieved in recent years both for *ab initio* and semiempirical quantum chemical methodology made possible to perform more or less precise calculations on truncated models representing some atoms of the enzyme active site interacting with the substrate (see e.g. [1]). Geometry optimization both in equilibrium and transition states became routine, as well, allowing to refine the molecular model further. However, it is still an open question to which extent the truncated model can be transferred from the gas phase to the enzyme, i.e. whether the optimized geometry obtained for this model neglecting the protein environment is a true representation of the real active site or not.

Correspondence to: G. Náray-Szabó

In our effort to treat the catalytic mechanism of aspartic proteases by theoretical methods we investigated the protonation state of the active site formed by a carboxyl–carboxylate dyad [2]. We tried to mimic the strain imposed by the protein environment on the dyad geometry by fixing carbon atoms at their experimental positions. It is questionable whether this freezing of the geometry is necessary at all.

Starting from the above problem we performed a detailed study on the transferability of the –COOH ··· $^{-}$OOC– geometry from the gas phase to crystals and proteins, as well. This is also interesting from another point of view, namely the appropriate comparison of quantum chemically calculated hydrogen-bonded O ··· O distances with experiment. As it is known data for hydrogen-bonded gas-phase complexes are not readily available [3] therefore selection of reference geometries, obtained for molecular crystals containing this entity, are important. Similar studies on covalent bond dihedral angles in substituted phenylbenzoate derivatives [4] and hydrogen-bonded geometry of hydrated imidazole derivatives [5] have shown that transferability is fair allowing to use experimental crystal structure data as reference for quantum chemical calculations in the gas phase. It is also worth for mentioning that Kanters et al. also analyzed the geometry of the carboxyl–carboxylate moiety [6] but they focused on covalent bond lengths and angles and not on H-bonded distances. Our study is an extension of the work by Sawyer and James, who compared the geometry of the carboxyl–carboxylate moiety for 4 crystalline associations and 4 proteins [7] and observed a certain transferability within this group.

Models and methods

We did *ab initio* and semiempirical Hartree–Fock molecular orbital calculations on the simplest model of the carboxyl–carboxylate entity, the HCOOH ··· $^{-}$OOCH system. We used the TEXAS program [8] and a 6-31G**$^{++}$ basis set including polarization and diffuse functions both on hydrogen and non-hydrogen atoms. This is necessary to describe correctly the negative charge distribution located near the carboxylate moiety. For the semiempirical calculation we selected the MNDO/PM3 method [9] that is known to provide fair geometries for most hydrogen-bonded systems and used the MOPAC program [10] for the calculation.

All four possible conformations (cf. Fig. 1) were treated separately with geometries fully optimized. In order to reduce the computer time we calculated torsional energy curves only by the MNDO/PM3 method. Starting from the models of Fig. 1, we fixed the selected dihedral angle and optimized all other geometry parameters. In addition, we also considered symmetric structures where the O3–H4 and H4–O6 distances were kept equal. In order to clarify the effect of bulky substituents on the relative stability of various conformers we studied the $(CH_3)_2$ CHCOOH ··· $^{-}$OOCCH$(CH_3)_2$ system by the PM3 method, as well. The convergence criteria in the *ab initio* calculation were: 10^{-9} a.u. for the SCF iteration limit and 10^{-6} a.u. Å^{-1} for the maximum force on each atom. The corresponding limits in the PM3 calculation were 10^{-6} a.u. for the SCF iteration limit and 10^{-3} kcal/mol pm for the maximum force on each atom.

We used the 1991 January files of the Cambridge Structural Database (CSD) [11] with more than 80000 total entries to select structures containing the –COO^{-} ··· HOOC– moiety. Bond length limits for the search were (cf. Fig. 1 for notation of atoms, in pm): C(H)1–C2: 145–155, C2–O5: 115–130, C2–O3:

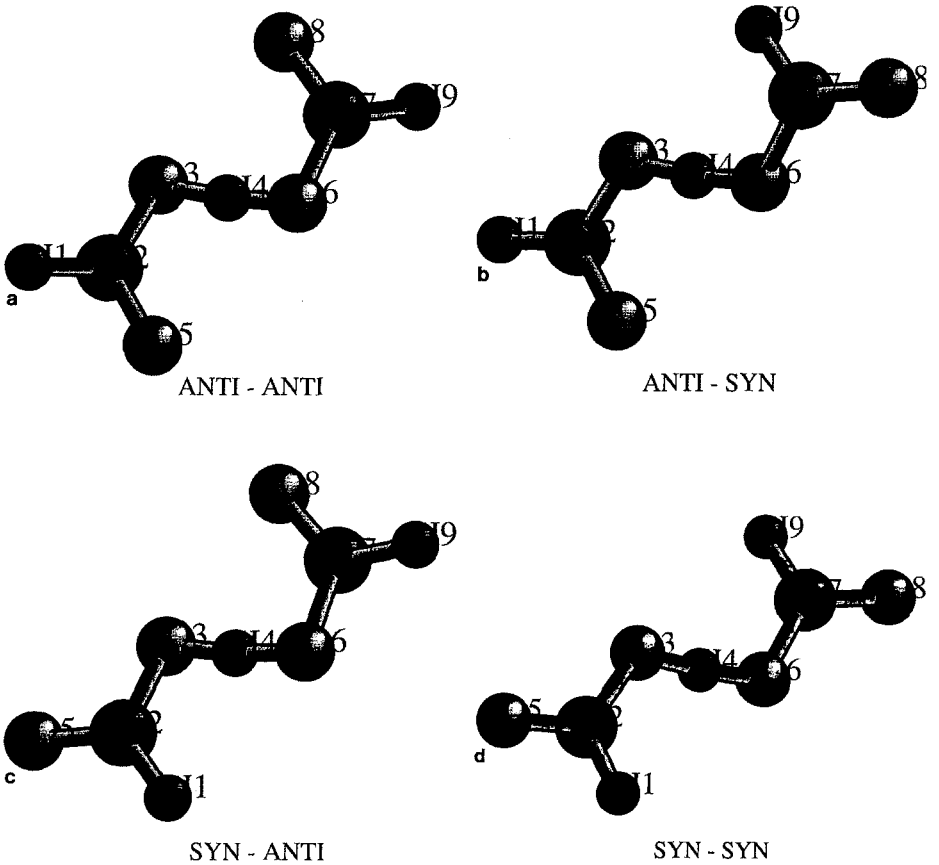

Fig. 1a–d. Possible conformations of the carboxyl–carboxylate couple. **a** *anti–anti* (τ_1, $\tau_3 > 150°$ or τ_1, $\tau_3 < -150°$), **b** *anti–syn* ($\tau_1 > 150°$ or $\tau_1 < -150°$ and $-30° < \tau_3 < 30°$), **c** *syn–anti* ($-30° < \tau_1 < 30°$ and $\tau_3 > 150°$ or $\tau_3 < -150°$), **d** *syn–syn* ($-30° < \tau_1$, $\tau_3 < 30°$) with $\tau_1 = O_6O_3C_2H_1 <$, $\tau_2 = C_7O_6O_3C_2 <$, $\tau_3 = H_9C_7O_6O_3 <$

122–130, O3–H4: 80–140, O6–H4: 80–210, O6–C7: 122–135, C7–O8: 115–130, C7–C(H)9: 145–155. For the symmetric structures the difference in O3–H4 and O6–H4 distances was less than 20 pm. Atomic groups that do not fit these criteria were dropped from our data set.

We applied the QUEST program package [12] with the following search profile

```
T1 *BTEST 88 + -92        AT1 O 1
T2 *CONNSER               AT2 C 3
AT1 O 1 O -1              AT3 O 1 1 E
AT2 C 3 O E               AT4 C 1
AT3 O 1 O E               BO 1 2 2
AT4 C 1                   BO 2 3 1
BO 1 2 1                  BO 2 4 1
BO 2 3 2                  END
```

```
BO 2 4 1                STOP 500
END                     SAVE 3 4
T3 *CONNSER             QUEST (T1). AND. (T2).AND (T3)
```

With this profile we located 238 entries representing hydrogen-bonded carboxyl–carboxylate moieties. Hydrogen bond lengths and relevant torsional angles were determined by the GEOM78 program. We defined the variable τ_1, τ_2, and τ_3 as the (6-3-2-1), (7-6-3-2) and (9-7-6-3) dihedral angles and defined four conformations depicted in Fig. 1. Among the 238 entries containing the $-COO^- \cdots HOOC-$ moiety we found 60 symmetric structures.

We located carboxyl–carboxylate pairs in the Summer 1992 Protein Data Bank files with a FORTRAN program that selects those pairs for which the non-bonded $O \cdots O$ distance is less than 360 pm. Since the hydrogen positions were not available we generated hydrogens on the carboxyl oxygens with standard geometric parameters ($r_{OH} = 96$ pm, COH $< = 104.5°$, CCOH $< = 0°$ and 180°) and included those structures in our data set where the $OH \cdots O$ non-bonded angle was larger than 150°. Defining the conformers like in case of the CSD search we located the following number of structures belonging to the various conformers: *anti–anti*: 21, *anti–syn*: 30, *syn–anti*: 19, *syn–syn*: 16.

Results and discussion

Transferability from the gas phase to crystals

In Table 1 we compared heats of formation of structures of Fig. 1 as calculated both by the *ab initio* and MNDO/PM3 methods in the gas phase to the number of hits found in CSD and corresponding to the given conformer. Larger heat of formation corresponds to a larger population of the given conformer in the gas phase. As we see both *ab initio* and semiempirical calculations predict the gas-phase *syn–anti* and *syn–syn* conformers to be more stable while in the crystal the *syn–syn* and *anti–syn* conformers are much stronger populated. The relative stability does not change in the gas phase even if we consider possible steric effects by putting a bulky isopropyl substituent at position 8. In Ref. [2] we have shown that the MNDO/PM3 method correctly reproduces conformational energy differences for the $HCOO^- \cdots HOH$ system as calculated by large basis set *ab initio* calculations including electron correlation. This result, combined with the agreement of *ab initio* and semiempirical calculations for the relative stability in Table 2 indicate that the conformers are not transferable from the gas phase to the crystal. The reason for this is most probably the effect of crystal packing.

In Fig. 2 the $O \cdots O$ distance histogram for all entries found in CSD are displayed. There is a marked maximum near 245 pm which is due to the presence of structures with symmetric H-bonds, i.e. those for which the difference between the O3H4 and O6H4 distances is less than 20 pm. If we treat these structures separately (cf. Fig. 3) the histograms will show closer similarity to a Gaussian distribution. Average $O \cdots O$ distances are 245 and 256 pm for the symmetric and asymmetric cases, respectively. Our *ab initio* calculations yield 236 pm for the symmetric and 253, 243, 255 and 261 pm for the asymmetric *anti–anti*, *anti–syn*, *syn–anti* and *syn–syn* structures, respectively. The weighted average, using the populations in Table 1, is 252 pm. On the other hand, MNDO/PM3 calculations give 225, 265, 263, 266 and 269 pm for these values, respectively. Our *ab initio*

Table 1. Fully optimized total energies (kcal/mol, upper row: *ab initio* – 236,500, lower row: PM3) and number of hits in the CSD search for conformations of Fig. 1. Lower indices refer to the group in position 8, iPr stands for isopropyl. Δ refers to the energy of the given conformation minus that of the *anti–anti* one

Conformation	Hits	$-E_H$	ΔE_H	$-E_{iPr}$	ΔE_{iPr}
anti–anti	122	17.5	0.0	–	–
		223.5	0.0	257.5	0.0
anti–syn	30	17.7	0.2	–	–
		222.5	– 1.0	256.8	– 0.7
syn–anti	3	19.7	2.2	–	–
		232.4	8.9	263.8	6.3
syn–syn	9	23.5	6.0	–	–
		227.0	3.5	263.1	5.6

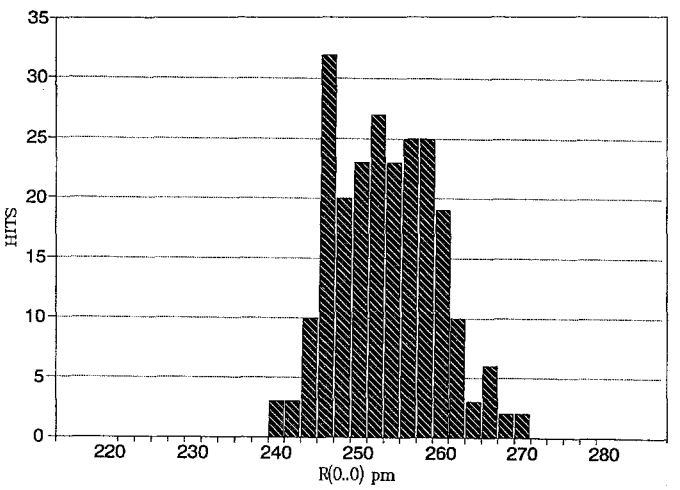

Fig. 2. Histogram showing the distribution of O ··· O distances in crystals including all four conformations of Fig. 1, symmetric and asymmetric arrangements, as well

calculations underestimate experimental O ··· O bond lengths by 6 pm in average that is somewhat more than the average error for hydrogen-bonded systems using a 6-31G* basis set (1–2 pm [13]). This means that in average a slight contraction in the bond length may take place upon transformation of the –COO⁻ ··· HOOC– moiety from the gas phase to the crystal but its magnitude is much smaller than that mentioned by Lesyng et al. (25 pm [14]). A maximum deviation of 7 and 14 pm from the mean value is observed in the histograms for the symmetric and asymmetric structures, respectively (considering at least 2 hits, cf. Fig. 3).

In order to give some explanation for the largest deviations from the mean value of the O ··· O distances we consider cases separately where this distance is much smaller or much larger than the average. An example for the former is norpropoxyphene maleate where the O ··· O distance obtained by X-ray diffraction is 240 pm [15] while optimisation of the molecular geometry *in vacuo* with the

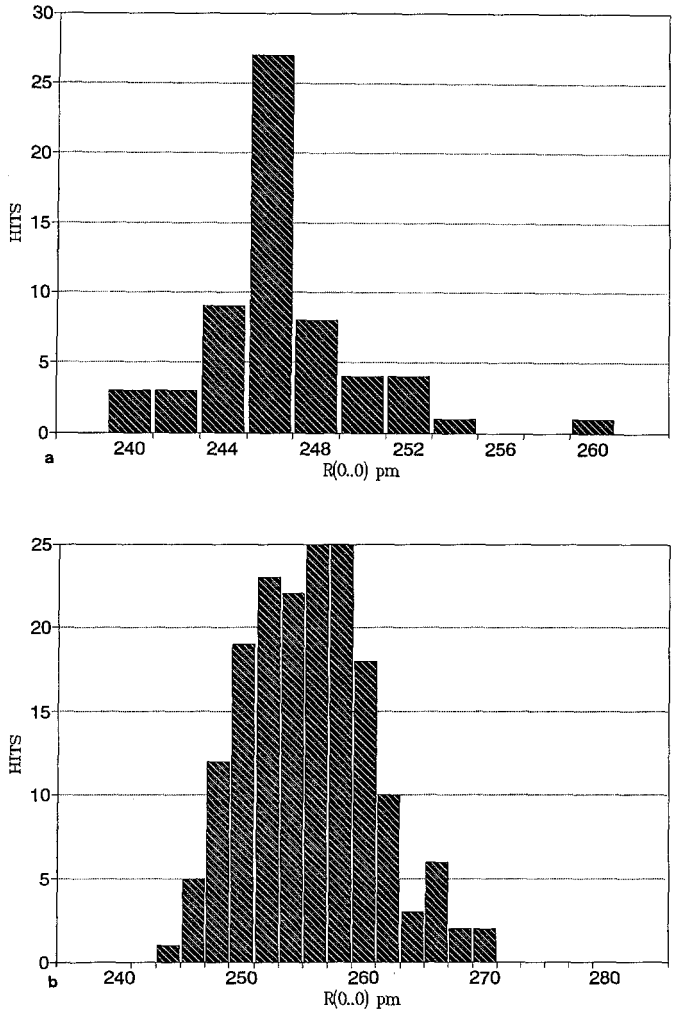

Fig. 3. The same as for Fig. 2 displaying symmetric (**a**) and asymmetric (**b**) H-bonds. For the symmetric H-bond the difference between O3H4 and O5H4 distances is less than 20 pm

PM3 method yields 255 pm. This means that for the free molecule no shortening is observed as compared to the average (256 pm). Reduction of the O ··· O distance in the crystal is most probably due to the close vicinity of a positive counter ion (protonated amine) that brings the carboxyl and carboxylate moieties, both hydrogen-bonded to the cation, closer to each other. O ··· O distances much longer than the average were found exclusively for intermolecular hydrogen bonds and in this case the elongation should be due to packing effects.

Figures 4–7 display histograms for the distribution of τ_2 in crystals for the *anti–anti*, *anti–syn*, *syn–anti* and *syn–syn* conformations. We compare these to the gas-phase PM3 torsional energy curve. For the *anti–anti* and *anti–syn* conformations (Figs. 4, 5) the maxima of the histograms are shifted by 50–60° relative to the

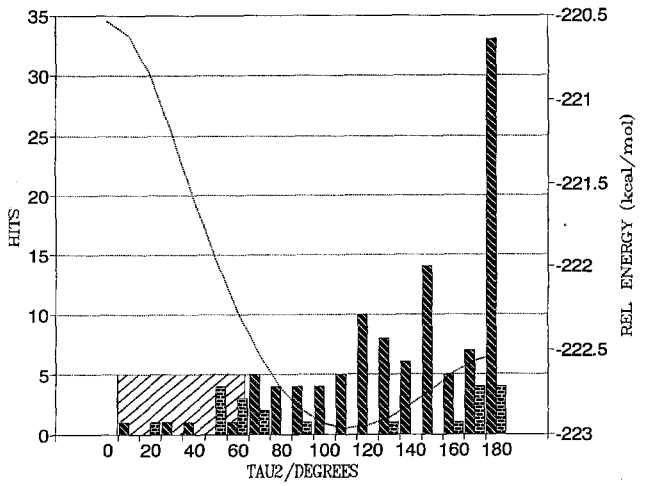

Fig. 4. Histogram showing the distribution of hits in CSD (cross-hatched bars) and PDB (bricks in bars) and the heat of formation as a function of the τ_2 torsional angle for the *anti–anti* conformation. Excluded regions of τ_2 are shaded

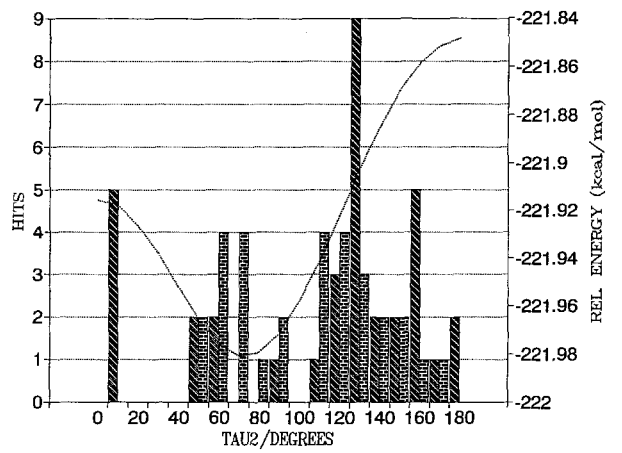

Fig. 5. The same as in Fig. 4 for the *anti–syn* conformation

minima of the semiempirical energy curves. The number of hits is only 3 for the *syn–anti* conformer (Fig. 6) therefore this case cannot be analyzed. On the other hand, for the *syn–syn* conformer (Fig. 7) the energy minimum and the maximum number of hits in the histogram coincide. In spite of the above disagreement the high energy regions of the torsional energy curves coincide almost perfectly with the excluded τ_2 range indicated by the histograms. In the 0–60° region for the *anti–anti* conformer (Fig. 4), where the torsional energy exceeds the minimum by kT, 0.6 kcal/mol at room temperature, showing that rotation is hindered here, we have only 4 hits, i.e. this range is excluded both in the gas and crystalline phases. For the *anti–syn* conformer the rotation energy change is very small, much beyond kT, so practically any value of τ_2 is energetically allowed which is reflected by the histogram, too (see Fig. 5). At last, for the *syn–syn* conformer no hits were found

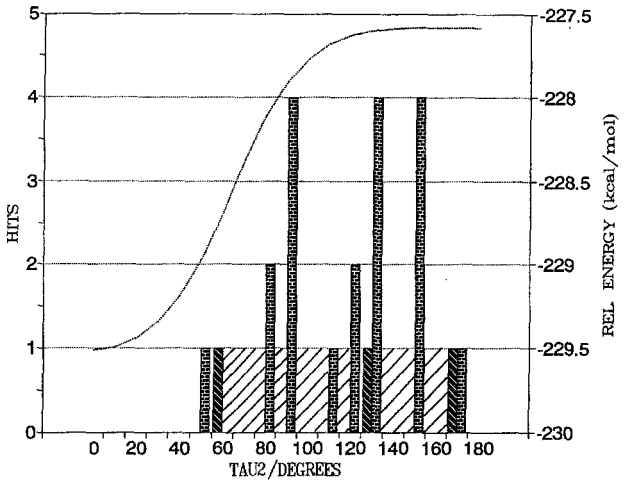

Fig. 6. The same as in Fig. 4 for the *syn–anti* conformation

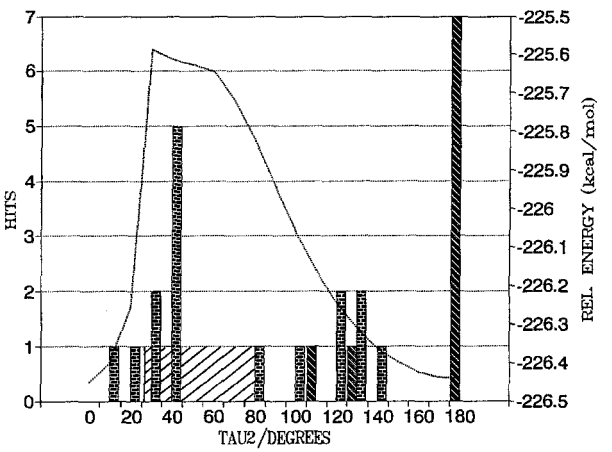

Fig. 7. The same as for Fig. 4 for the *syn–syn* conformation

in the 20–80° region where the rotational energy exceeds the minimum by 0.6 kcal/mol (Fig. 7). On the basis of the above considerations we may state that excluded regions for the torsional angle τ_2, as predicted by the MNDO/PM3 calculation, almost perfectly coincide with those indicated by the histograms depicting the experimental distribution in crystals. In a previous work we found a similar transferability of excluded values of torsional angles in substituted phenylbenzoates [4].

Transferability from the gas phase to proteins

Histograms of O···O distances in proteins from the PDB files are depicted in Fig. 8. The distribution for all hits is diffuse though it has a marked maximum between 265 and 270 pm (cf. Fig. 8a). If we consider only those hits that refer to one

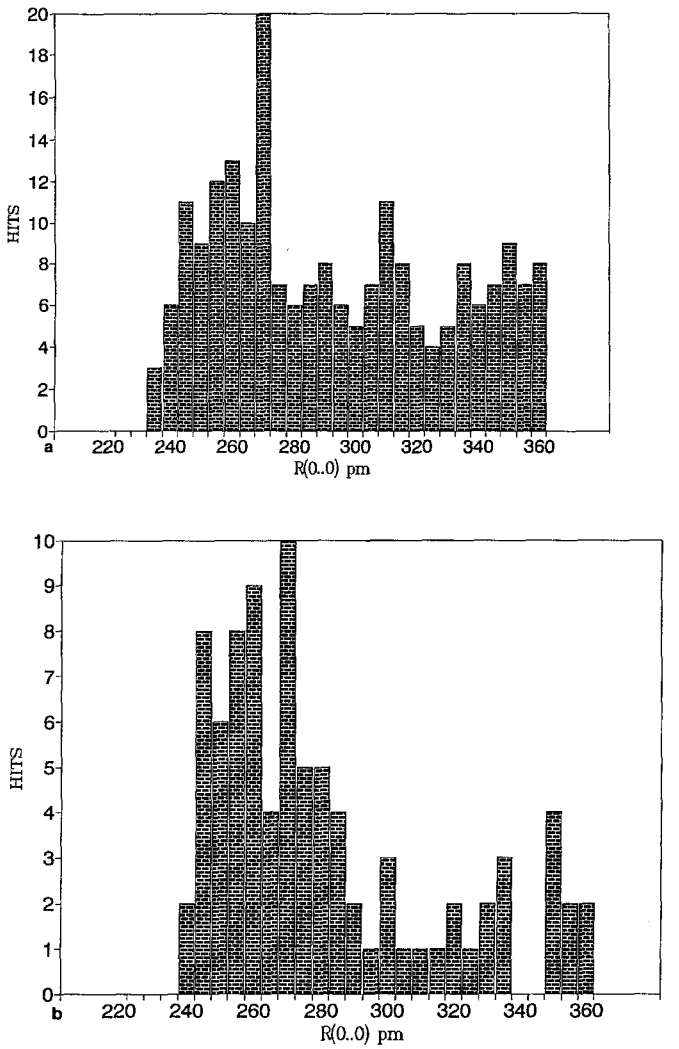

Fig. 8. Histogram showing the distribution of O···O distances in proteins. a) All hits in PDB, b) conformations displayed in Fig. 1

of the conformations in Fig. 1, O···O distances larger than 300 pm occur much less frequently (Fig. 8b). The average O···O separation for the latter case is 262 ± 27 pm, 10 pm larger than the *ab initio* value. The larger difference from the gas-phase value as well as the larger value of maximum deviation from the mean as compared to the crystalline phase is clearly due to the stronger constraint put on the geometry of the –OCC⁻···HOOC– moiety in proteins than in crystals.

In Figs. 5–7 we depict the distribution of τ_2 for the four different confor mations of the –COO⁻···HOOC– moiety embedded in proteins of the Summer 92 Protein Data Bank files. Histograms show no similarity to those obtained for

the Cambridge Structural Database entries. However, energetically excluded regions are avoided in proteins, too, even if less frequently than in crystals.

In their paper Sawyer and James mentioned 4 crystals and 4 proteins incorporating the $-COO^- \cdots HOOC-$ moiety [7]. $O \cdots O$ distances in the crystal vary between 237 and 253 pm (average 246 pm) and in proteins between 249 and 288 pm (average 263 pm). Most probably due to the small number of structures considered their average for crystals strongly differs from ours. However, for proteins the average $O \cdots O$ distance in their work is very close to that found by us. It is also clear from their study that the variation from the mean is larger for proteins than for crystals. As for the transferability of τ_2 values, they stated that "the values of torsion angles ··· for the proteins examined are not unusual when compared with those of acid salts" [7]. This is in agreement with our experience, too.

Conclusions

We found that upon transferring the $-COO^- \cdots HOOC-$ moiety from the gas phase to crystals and proteins (a) average $O \cdots O$ distances elongate to some extent, stronger in proteins than in crystals and (b) energetically excluded values for the τ_2 torsional angle appear only exceptionally in crystals and relatively rarely in proteins. We may state that the non-bonding geometry can be transferred from the gas phase to crystals and, with some caution, to proteins.

Acknowledgements. This work was supported by grants from the National Foundation for Scientific Research (OTKA) No. T4316 and Catching up with European Higher Education (FEFA) No. II/265.

References

1. Daggett V, Schröder S, Kollman P (1991) J Am Chem Soc 113:8926
2. Turi L, Náray-Szabó G (1992) Int J Quant Chem 42:1537
3. Legon AC, Millen DJ (1992) Chem Soc Rev 21:71
4. Birner P, Kugler S, Simon K, Náray-Szabó G (1982) Mol Cryst Liqu Cryst 80:11
5. Alagona G, Ghio C, Nagy P, Simon K, Náray-Szabó G (1990) J Comput Chem 11:1038
6. Kanters JA, Kroon J, Hooft R, Schouten A, Schijndel JAM, Brandsen J (1991) Croat Chem Acta 64:353
7. Sawyer L, James MNG (1982) Nature 295:79
8. Pulay P (1979) Theoret Chim Acta 50:229
9. Stewart JJP (1989) J Comput Chem 10:209
10. Merz KM Jr, Besler JH, MOPAC version 5.0, Quantum Chemistry Program Exchange, No. 589
11. Allen FH, Kennard O, Taylor R (1983) Acc Chem Res 16:146
12. Hehre WJ, Radom L, Schleyer PvR, Pople JA (1986) *Ab initio* molecular orbital theory. Wiley, New York
13. Lesyng B, Jeffrey G, Maluszynska H (1988) Acta Cryst B44:193
14. Taylor R, Kennard O (1982) J Am Chem Soc 104:5063
15. Bye E (1977) Acta Chem Scand Ser B31:157