Acta Cryst. (1991). A**47**, 842-846

A Multisolution Method of Phase Determination by Combined Maximization of Entropy and Likelihood. V. The Use of Likelihood as a Discriminator of Phase Sets Produced by the *SAYTAN* Program for a Small Protein

By C. J. Gilmore and A. N. Henderson

Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, Scotland

and G. Bricogne*

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England

(*Received* 22 *October* 1990; *accepted* 2 *May* 1991)

Abstract

The use of a likelihood criterion associated with maximum-entropy (ME) extrapolation for selecting phase sets as part of a new multisolution phasing strategy, already applied to solving small crystal structures from single-crystal data [Gilmore, Bricogne & Bannister (1990). *Acta Cryst.* A**46**, 297-308] and X-ray powder diffraction data [Gilmore, Henderson & Bricogne (1991). *Acta Cryst.* A**47**, 830-841], has been tested on the small protein avian pancreatic polypeptide (APP) with 301 non-H atoms in the asymmetric unit in space group $C2$. A collection of 50 phase sets for APP were provided by Woolfson & Yao. They had been generated from random starting phases by the *SAYTAN* procedure [Woolfson & Yao (1990). *Acta Cryst.* A**46**, 409-413] using data to a resolution of 0·98 Å. Six of these had an unweighted mean absolute phase error, $\langle|\Delta\varphi|\rangle$, of less than 50°, the remainder having phase errors of 60° or more. However, none of the conventional figures of merit were able to identify these preferred sets. Each phase set was subjected to our standard procedure of entropy maximization and of evaluation of the log-likelihood gain resulting from the associated ME extrapolation. With only a small subset of data (to 2 Å resolution), the likelihood criterion identified unambiguously the phase sets with $\langle|\Delta\varphi|\rangle$ less than 50°. In contrast, conventional figures of merit showed no such ability. We conclude that, although *SAYTAN* is at present much more efficient than our program (*MICE*) at generating large trial phase sets, the latter is clearly superior in discerning the best of these phase sets. We therefore expect that our current efforts to make *MICE* run faster should result in better prospects of *ab initio* phasing for small macromolecules.

0. Introduction

In previous papers (Gilmore, Bricogne & Bannister, 1990; Gilmore, Henderson & Bricogne 1991) we have applied the methods of entropy maximization and likelihood estimation to the phasing of small molecules from both single-crystal and powder diffraction data. In particular, likelihood emerged as a figure of merit of unique power in the selection of phase sets having minimum phase error.

Debaerdemaeker, Tate & Woolfson (1985, 1988) have devised an alternative formalism, programmed as *SAYTAN*, that replaces the traditional tangent formula of direct methods. The latter is a corner stone of direct methods where it is used both to generate and to refine phases. Since it treats all invariants as independent without taking account of their correlations, it tends to refine phases to overconsistency; it also requires data at atomic resolution. In *SAYTAN* (Sayre tangent formula) these problems are, in part, overcome by using not only the largest E magnitudes but also the smallest, and an attempt is made to satisfy the Sayre equation (Sayre, 1952) by the suitable incorporation of both positive and negative quartets and Ψ_0-type triplets, which involve two strong and one weak E magnitudes. It has been applied successfully to a small protein, avian pancreatic polypeptide (APP) (Woolfson & Yao, 1990) where 1000 phase sets were generated; the initial phases for the strongest 727 E magnitudes were assigned random values and these were refined by the Sayre tangent equation. Six of these 1000 sets had an unweighted mean absolute phase error, $\langle|\Delta\varphi|\rangle$, below 45° and thus contained useful structural information. However, there was considerable difficulty in distinguishing these sets using conventional figures of merit (FOM's). Since then a new criterion termed TFOM (Tate, private communication) has been developed which can act as a partial filter, reducing the 1000 sets to *ca* 20, but it still cannot wholly distinguish the best ones. Thus the time-consuming step of map generation and interpretation is necessary in order to apply this method to APP.

It was thought that this situation would provide a good test of the discrimination power of likelihood for small macromolecules. A large number of phase sets generated by *SAYTAN* were subjected to entropy

---

* Permanent address: LURE, Université Paris-Sud, 91405 Orsay, France.

0108-7673/91/060842-05$03.00

maximization using the program *MICE* (Gilmore, Bricogne & Bannister, 1990). The variation of the log-likelihood gain in the course of entropy maximization was recorded for each phase set to test our expectation that the best phase sets would exhibit large log-likelihood gains during this process.

§ 1 outlines data preparation, the initial *SAYTAN* phasing and the interface to the *MICE* program. § 2 presents and discusses the results in terms of likelihood gains, Σ refinement and the resulting maps. § 3 summarizes the results and outlines forthcoming further development.

1. Data preparation, *SAYTAN* phasing and the *MICE* program

Avian pancreatic polypeptide (Glover, Haneef, Pitts, Wood, Moss, Tickle & Blundell, 1983) is a small protein with the following crystal data: $C_{109}N_{53}O_{55}Zn + 80H_2O$, monoclinic, $a = 34{\cdot}18$, $b = 32{\cdot}92$, $c = 28{\cdot}40$ Å, $\beta = 105{\cdot}6°$; space group $C2$, $Z = 4$. There are 301 non-H atoms in 36 residues in the asymmetric unit. The data resolution is 0·98 Å.

All the data sets were kindly provided by Yao Jia-xing and Michael Woolfson. They had previously been processed as follows.

(i) The data were normalized to give E magnitudes. The largest 800 and the smallest 200 magnitudes were selected for use in *SAYTAN*. A convergence map rejected 33 large and 17 small E magnitudes on the grounds of insufficient triplet connectivity; the remaining 727 large E's interconnected by 9726 triplets and 183 small E's interconnected by 6434 Ψ_0 relationships were used in phase refinement.

(ii) Starting with random phases 1000 trials were made in which the phases were refined to convergence by *SAYTAN* employing a limited subset of positive quartets, negative quartets, three-phase invariants and Ψ_0 triplets. Three subsets of these 1000 sets were used as input to *MICE*:

(*a*) a set of 10 containing one good set. This acted as a pilot test to measure the usefulness of traditional figures of merit;

(*b*) a set of 50 containing six sets having $\langle|\Delta\varphi|\rangle < 50°$;

(*c*) a set of 20 filtered by the TFOM figure of merit.

It was not possible to process all 1000 phase sets because of the considerable demands such a calculation would have made on our computing resources.

Each of these phase sets were processed as follows.

(iii) A full set of E magnitudes was used: the likelihood explores how well structure-factor magnitudes are extrapolated for *all* intensities - not just the very strong and the very weak. No standard deviations were available for $|E_{\mathbf{h}}|$; these are useful in both the entropy maximization itself and in the likelihood calculations where they weight the contributions of individual reflections. They were simulated by setting $\sigma(|E_{\mathbf{h}}|) = 0{\cdot}1|E_{\mathbf{h}}| + 0{\cdot}01$. This corresponds to the normal level of *ca* 10–15% error in E magnitude found in good-quality data sets. Since the variance of $|E_{\mathbf{h}}|$ plays only a minor role in entropy maximization and since the procedure is not unduly sensitive to it, this relatively simple method does not invalidate or bias the subsequent calculations in any way. The E magnitudes were converted to unitary structure factors, $|U_{\mathbf{h}}|^{\text{obs}}$.

(iv) For each set of phases, the data were split into two: a set $\{H\}$, the basis set, of phased reflections from *SAYTAN* which were used as constraints in entropy maximization to generate the non-uniform prior $q^{\text{ME}}(\mathbf{x})$ and the remainder $\{K\}$ which were employed in likelihood estimation. The *MICE* program uses a modified method of exponential modelling (Collins & Mahar, 1983; Bricogne, 1984). In order to overcome the inherent instability of the naive algorithm, it is modified to include two search directions, the second of which serves to temper the build up of contrast in $q^{\text{ME}}(\mathbf{x})$ (Bricogne & Gilmore, 1990). Some of the steps of this method are most efficiently carried out in real space, and thus each cycle of entropy maximization necessitates the computation of up to 14 Fourier transforms. To reduce aliasing errors, these maps must be oversampled and a grid size of less than 0·2 Å is necessary when processing data at 0·98 Å resolution (which is the resolution of the APP data set). Although this is possible, it has the consequence of being very slow to compute. Thus, to save time, only those phased reflections having a resolution of 2·0 Å or less were incorporated into the basis set. There were 117 of these.

(v) The 117 basis-set reflections ($\mathbf{h} \in H$) were used as constraints in the generation of a maximum-entropy prior $q^{\text{ME}}(\mathbf{x})$. The latter reproduces the known phases and amplitudes for reflections $\mathbf{h} \in H$, but remains maximally non-committal with respect to the unknown structure factors $\mathbf{h} \in K$. Thus the construction of $q^{\text{ME}}(\mathbf{x})$ gives rise to phase extrapolation, *i.e.* its Fourier transform can generate $|U_{\mathbf{h}}^{\text{ME}}|$ and $\varphi_{\mathbf{h}}^{\text{ME}}$ for both $\mathbf{h} \in H$ and $\mathbf{h} \in K$.

Likelihood measures the fit between $|U_{\mathbf{h}}|^{\text{obs}}$ and $|U_{\mathbf{h}}^{\text{ME}}|$ for $\mathbf{h} \in K$ taking into account the probability distribution of $\varphi_{\mathbf{h}}$. It was monitored from cycle to cycle for each phase set. The diagonal approximation [Bricogne (1984); Bricogne & Gilmore (1990), § 2.4] was used throughout. A null hypothesis, $\mathcal{H}_0$, is first defined by substituting $|U_{\mathbf{h}}^{\text{ME}}| = 0$ into the log-likelihood expressions and summing over all extrapolates, with separate terms for centric and acentric reflections. This gives a log-likelihood $L\mathcal{H}_0$. The extrapolated U magnitudes, $|U_{\mathbf{h}}^{\text{ME}}|$, are now substituted to define the hypothesis $\mathcal{H}$, giving a log-likelihood value $L\mathcal{H}$. The difference $L\mathcal{H} - L\mathcal{H}_0$ is the log-likelihood gain and acts as a figure of merit, although its maximization can also be used for phase refinement [Bricogne & Gilmore (1990) § 2.5].

Table 1. *A comparison of maximum log-likelihood gain and traditional figures of merit* (Ψ_0, R_α *and ABSFOM*) *for ten SAYTAN-derived phase sets*

All phase errors are in degrees; χ^2 is the χ^2 statistic at the likelihood maximum. (For correct phase sets Ψ_0 should be close to unity, R_α a minimum and ABSFOM in the range 1·0–1·2.)

Set no.	Absolute phase error	Maximum $L\mathscr{H}-L\mathscr{H}_0$	χ^2	Ψ_0	R_α	ABSFOM
1	79·7	−0·01	2·47	0·87	26·95	0·66
2	39·1	11·26	1·81	2·73	29·24	1·70
3	84·5	0·19	1·58	0·77	27·21	0·67
4	82·9	0·07	2·25	0·84	26·33	0·69
5	84·8	−0·02	3·44	0·90	27·53	0·70
6	84·3	0·00	2·52	0·83	26·94	0·67
7	84·5	0·36	1·40	0·89	26·18	0·72
8	81·0	3·47	0·97	0·82	26·04	0·72
9	83·3	0·01	2·98	0·90	26·15	0·70
10	84·9	0·09	2·45	0·80	27·24	0·69

(iv) The fit between $U_{\mathbf{h}}^{\rm obs}$ and $U_{\mathbf{h}}^{\rm ME}$ for $\mathbf{h}\in H$ is of great importance. As a measure of fit we used the reduced-χ^2 statistic:

$$\chi^2=(2n_a+n_c)^{-1}\sum_{\mathbf{h}\in H} s_{\mathbf{h}}^2\|U_{\mathbf{h}}|^{\rm obs}-|U_{\mathbf{h}}|^{\rm ME}|^2.$$

n_a and n_c are the numbers of unique acentric and centric reflections in the basis set respectively. The sum $2n_a+n_c$ is the total number of degrees of freedom. The parameter $s_{\mathbf{h}}$ is a measure of variance and has four components:

$$s_{\mathbf{h}}^2=\sigma_{\mathbf{h}}^2+p\varepsilon_{\mathbf{h}}\Sigma$$

where $\varepsilon_{\mathbf{h}}$ is the standard epsilon factor, $\sigma_{\mathbf{h}}^2$ is the estimated variance of $|U_{\mathbf{h}}|^{\rm obs}$. Σ is obtained from likelihood refinement and is related to the structural complexity of the unit cell and the data resolution of the basis-set reflection, whilst p is an empirical parameter set to unity throughout this work. Two Σ parameters are refined: one based on the acentric extrapolates Σ_a and a corresponding centric parameter Σ_c. The effective number of atoms in the unit cell, $N_{\rm eff}$, is given as the weighted mean of $1/\Sigma_c+2/\Sigma_a$. A newer likelihood formalism (Bricogne, 1991) employs a single Σ parameter and is thus simpler, but was not used here because this research was carried out before the programming of the new method was completed.

Entropy maximization continued to $\chi^2=1\cdot0$. In general, the likelihood passed through a maximum before this point, but it was considered important to monitor all likelihood variations for each phase set and so refinement was not terminated at the point of maximum likelihood.

2. Results and discussion

Table 1 shows a comparison between maximum log-likelihood gain and the traditional figures of merit Ψ_0, R_α and ABSFOM for ten *SAYTAN*-derived phase sets. The figures of merit were calculated using the *MITHRIL* direct-methods program (Gilmore, 1984; Gilmore & Brown, 1988). Nine of these sets have $\langle|\Delta\varphi|\rangle$ greater than 80° and the log-likelihood gains vary from −0·02 to 3·47. The remaining set, no. 2, has an error of only 39·1° with a corresponding log-likelihood gain of 11·26; it is thus clearly indicated. As previously reported by Woolfson & Yao, none of the other figures of merit give any useful indication; indeed, the best set has by far the worst Ψ_0 value, *i.e.* the highest; the R_α indicator exhibits a variation of less than 12%, and it too has a maximum value (*i.e.* the worst) for the correct set. The ABSFOM value of 1·70 is much too large and such a solution would be rejected on the grounds of overconsistency of contributing invariants. The negative quartet figure of merit, NQEST (De Titta, Edmonds, Langs & Hauptman, 1975) was not used since the negative quartets were deemed too few and too unreliable by the *MITHRIL* program.

The behaviour of likelihood as a function of χ^2 in these calculations is atypical. In previous calculations we have found a regular increase in likelihood gain from its starting value as entropy maximization proceeds. Some sets, *e.g.* no. 8, do show this, but in general the behaviour is different and most sets have likelihood maxima in the range $1\cdot5<\chi^2<3\cdot5$. Correct sets tend to have maxima at $\chi^2\simeq1\cdot7$. This behaviour is most probably a consequence of the Zn atoms which form a pseudocentrosymmetric array in the cell and dominate the X-ray scattering. Indeed, all the centroid maps at 2·0 Å derived from these *SAYTAN* phase sets display only the Zn atoms, leaving a problem of enantiomorph definition and pseudosymmetry when completing the structure with standard crystallographic methods. It is significant that the existence of residual ambiguities in the *SAYTAN* phase sets was indicated, perhaps somewhat cryptically, by an unusual behaviour of the likelihood criterion. We would expect that in the tree-directed multisolution strategy described elsewhere, this persistent enantiomorph ambiguity would have been detected at an early stage and would have given rise to the growth of two parallel branches (one for each of the possible enantiomorph choices) until enough phase information has accrued for likelihood to discriminate between them. Along each of these branches, likelihood would increase monotonically for the best node as χ^2 decreased to 1·0. These predictions will be tested when we attempt the *ab initio* phasing of APP with *MICE*.

Because different sets have different χ^2 values, it is not worthwhile to tabulate entropies, since the latter is a function of the χ^2 statistic and becomes increasingly negative as χ^2 decreases.

Table 2 summarizes a similar calculation for a collection of 50 phase sets produced by *SAYTAN*. There are six sets present with $\langle|\Delta\varphi|\rangle$ less than 50° and they are all readily identified by the likelihood criterion. This is a remarkable result; only the 117

Table 2. *Log-likelihood gain as a figure of merit for 50 phase sets produced by SAYTAN*

All phase errors are in degrees. $L\mathscr{H}-L\mathscr{H}_0$ is the maximum log-likelihood gain; the corresponding χ^2 is tabulated in column 5; Σ_a and Σ_c are discussed in the text.

Set no.	R.m.s. Phase error	Mean absolute phase error	$L\mathscr{H}-L\mathscr{H}_0$	χ^2	$\Sigma_a\times10^3$	$\Sigma_c\times10^3$
1	100·6	85·4	3·59	0·98	0·66	0·96
2	100·4	84·4	1·36	1·22	0·67	0·96
3	100·3	84·7	0·07	2·41	0·68	0·95
4	99·0	82·5	1·92	0·98	0·66	0·95
5	99·2	83·7	0·33	1·49	0·67	0·97
6	99·4	83·3	2·61	1·25	0·67	0·95
7	59·7	42·7	10·52	1·93	0·62	0·89
8	99·9	84·3	−0·01	3·07	0·69	1·06
9	100·9	85·6	0·14	1·77	0·68	0·99
10	100·5	85·0	0·22	1·51	0·67	0·96
11	63·3	47·1	11·10	1·93	0·62	0·89
12	100·4	85·0	0·18	2·05	0·68	0·99
13	99·7	83·5	0·45	1·53	0·67	0·97
14	96·8	80·2	3·60	0·97	0·65	0·92
15	64·6	48·6	8·47	2·03	0·63	0·91
16	100·0	85·0	1·28	1·58	0·66	0·95
17	100·0	84·1	0·05	3·05	0·69	1·03
18	64·9	48·1	10·49	1·95	0·62	0·88
19	101·3	85·6	0·00	3·15	0·69	0·98
20	97·8	82·0	1·90	1·52	0·66	0·95
21	101·1	85·7	1·80	0·99	0·66	0·94
22	65·8	49·2	4·58	2·35	0·66	0·95
23	99·3	84·0	0·17	2·04	0·68	0·98
24	101·5	85·9	1·85	1·30	0·66	0·95
25	99·4	83·3	1·09	0·98	0·66	0·94
26	101·1	85·4	2·09	1·30	0·66	0·94
27	97·8	82·2	0·09	2·81	0·68	1·02
28	98·7	83·0	−56·68	1·80	0·68	0·97
29	101·8	85·8	0·48	1·99	0·67	0·96
30	100·5	84·8	0·08	2·62	0·68	1·01
31	96·8	80·6	−0·02	3·30	0·70	1·12
32	58·6	43·2	10·56	1·94	0·63	0·89
33	100·4	85·1	0·11	1·49	0·67	0·99
34	100·2	84·5	0·08	1·76	0·68	0·97
35	99·0	83·3	−0·02	3·20	0·69	1·06
36	100·8	85·5	0·07	1·78	0·68	0·99
37	101·0	85·2	0·28	1·54	0·67	0·98
38	98·6	82·4	3·42	0·98	0·66	0·93
39	100·4	84·7	0·02	2·61	0·68	0·98
40	100·9	85·4	0·01	1·74	0·68	0·99
41	100·5	85·4	−4·79	3·43	0·70	1·14
42	98·6	83·2	−15·92	2·83	0·69	1·04
43	98·1	82·1	−41·97	3·28	0·68	1·00
44	96·8	81·1	−18·55	3·48	0·69	1·12
45	100·4	83·9	−31·61	3·35	0·70	1·12
46	99·7	84·1	−0·04	3·10	0·69	1·07
47	100·6	85·4	0·12	1·82	0·68	0·97
48	100·0	84·6	2·97	0·97	0·66	0·95
49	101·7	86·1	0·68	1·55	0·67	0·98
50	98·2	81·9	0·01	3·37	0·71	1·08

Table 3. *Log-likelihood gain for ten phase sets from SAYTAN*

In this case, *SAYTAN* produced 1000 sets and these were first filtered by the TFOM figure of merit before being passed to the *MICE* program. The column headings are as for Table 2.

Set no.	R.m.s. phase error	Mean absolute phase error	$L\mathscr{H}-L\mathscr{H}_0$	χ^2	$\Sigma_a\times10^3$	$\Sigma_a\times10^3$
1	97·76	80·29	0·11	3·40	0·68	1·01
2	97·62	80·55	0·45	1·39	0·67	1·04
3	94·56	76·83	0·06	2·87	0·68	1·05
4	93·90	76·55	0·08	3·41	0·68	1·01
5	94·33	76·56	1·31	2·46	0·67	1·00
6	90·78	71·77	0·03	3·22	0·69	1·14
7	67·48	49·46	10·01	2·00	0·63	0·92
8	64·69	46·11	10·01	1·73	0·58	0·85
9	96·21	79·06	0·07	3·57	0·68	1·01
10	56·17	40·59	10·56	1·79	0·60	0·88
11	95·09	76·50	0·05	3·57	0·68	1·01
12	56·78	41·10	11·89	1·83	0·61	0·89
13	92·96	74·23	−0·00	3·32	0·70	1·22
14	98·13	80·61	0·07	3·42	0·68	1·01
15	98·12	80·16	10·22	2·01	0·63	0·92
16	92·90	74·37	0·09	2·68	0·68	1·06
17	99·70	82·27	0·06	3·58	0·68	1·01
18	88·83	70·15	1·08	1·41	0·66	0·97
19	92·61	75·52	0·56	1·31	0·66	0·99
20	94·57	76·98	0·53	1·61	0·67	0·99

largest *U* magnitudes at 2·0 Å resolution have been input into the basis set, and yet the power of discrimination is excellent. Set 22 is rather more weakly indicated than the others, but it was stopped at a χ^2 of 2·35, and thus is considerably underfitted with correspondingly weak extrapolation.

The stability of entropy maximization at this resolution is also clearly demonstrated. The refined Σ_a and Σ_c values are *ca* $6{\cdot}6\times10^{-4}$ and $0{\cdot}96\times10^{-4}$ except for the best phase sets which have lower values of about $6{\cdot}2\times10^{-4}$ and $0{\cdot}90\times10^{-4}$. The weighted mean of these latter values gives an $N_{\rm eff}$ value of 2620. The true N is *ca* 1600; this larger value of $N_{\rm eff}$ reflects the accumulation of a good deal of correct phase information with correspondingly exact phase extrapolation [Bricogne & Gilmore (1990), § 3.3]. It would, however, be unwise to use a minimum Σ criterion as a figure of merit.

Table 3 summarizes the results pertaining to a collection of 20 phase sets selected by the TFOM figure of merit out of 1000 trial phase sets produced by *SAYTAN*. Again, the correct sets are indicated, but with one anomaly - set 15, which is probably the result of a local entropy maximum. The corresponding map is uninterpretable: even the Zn atoms cannot be located and so this solution would probably be rejected at higher resolution.

Finally, it is worth recording that APP has been solved *ab initio* by G. Sheldrick (private communication). As in this work, the Zn atoms were readily located and the structure was completed using the sophisticated fragment recycling procedures of *SHELXS* (Sheldrick, 1990).

3. Concluding remarks

We have shown once again that likelihood, coupled with entropy maximization, is a uniquely powerful discriminator of correct phase sets. In this case it was able to distinguish correctly those solutions with a mean absolute phase error of less than 50° even when employing only 117 reflections to a resolution of 2 Å in the basis set.

In this work, we have not attempted to phase APP *ab initio*, but we have instead used phase sets generated by the *SAYTAN* procedure. Clearly, *SAYTAN* is much more efficient at generating large trial phase sets than our present version of *MICE* but much poorer at recognizing the best of them. We are currently reprogramming the calculation and handling of maps in *MICE* to eliminate all disk transfers, which should lead to a speed up by a factor of five. We expect this will bring the phase-extension capabilities of *MICE* up to those of *SAYTAN*; the superiority of likelihood over other figures of merit should then confer a definite advantage to the maximum-entropy approach over other direct methods. We will put these expectations to the test by attempting the direct structure determination of APP as soon as possible.

We are grateful to Professor Michael Woolfson and Dr Yao for providing the *SAYTAN* phase sets and much useful discussion. This research was supported by an SERC grant. GB acknowledges the unstinting support of Trinity College, Cambridge.

References

Bricogne, G. (1984). *Acta Cryst.* **A40**, 410-445.
Bricogne, G. (1991). *Acta Cryst.* **A47**, 803-829.
Bricogne, G. & Gilmore, C. J. (1990). *Acta Cryst.* **A46**, 284-297.
Collins, D. M. & Mahar, M. C. (1983). *Acta Cryst.* **A39**, 252-256.
Debaerdemaeker, T., Tate, C. & Woolfson, M. M. (1985). *Acta Cryst.* **A39**, 286-290.
Debaerdemaeker, T., Tate, C. & Woolfson, M. M. (1988). *Acta Cryst.* **A44**, 353-357.
De Titta, G. T., Edmonds, J. W., Langs, D. A. & Hauptman, H. (1975). *Acta Cryst.* **A31**, 472-479.
Gilmore, C. J. (1984). *J. Appl. Cryst.* **17**, 42-46.
Gilmore, C. J., Bricogne, G. & Bannister, C. (1990). *Acta Cryst.* **A46**, 297-308.
Gilmore, C. J. & Brown, S. R. (1988). *J. Appl. Cryst.* **22**, 571-572.
Gilmore, C. J., Henderson, K. & Bricogne, G. (1991). *Acta Cryst.* **A47**, 830-841.
Glover, I., Haneef, I., Pitts, J.-E., Wood, S. P., Moss, D., Tickle, I. J. & Blundell, T. L. (1983). *Biopolymers*, **22**, 293-304.
Sayre, D. (1952). *Acta Cryst.* **5**, 60-65.
Sheldrick, G. M. (1990). *Acta Cryst.* **A46**, 467-473.
Woolfson, M. M. & Yao, J. (1990). *Acta Cryst.* **A46**, 409-413.

Acta Cryst. (1991). A47, 846-850

Elastic-Tensor Atom–Atom Potential Calculations for Molecular Crystals: C_6H_6 and $CO(NH_2)_2$

By P. Pavlides, D. Pugh and K. J. Roberts*

University of Strathclyde, Department of Pure & Applied Chemistry, Glasgow G1 1XL, Scotland

(*Received* 13 *August* 1990; *accepted* 21 *June* 1991)

Abstract

The numerical method for calculating the elastic constants of molecular crystals, using the crystal packing program *PCK*83, developed and applied to the model case of CO_2 in previous work is here extended to benzene (C_6H_6) and urea [$(NH_2)_2CO$] crystals. Interatomic potential-energy functions coupled with rigid-molecule constraints are used, taking the entire angular dependence into account. Correlation with other calculations and experimental data is good.

Introduction

Various methodologies for modelling intermolecular and interatomic forces in crystals have been developed over the years, in crystal packing programs such as *PCK*83 (Williams, 1983) and *WMIN* (Busing, 1981), with a view to predicting their physicochemical properties (Catlow, Cox, Jackson, Parker, Price, Tomlinson & Vetrivet, 1989; Kitaigorodsky, 1973, 1978). Of particular interest, and the subject of this paper, are the elastic constants of crystals, which are related to the second derivatives of the energy hypersurface at the minimum. Knowledge of the elastic behaviour of solids is of importance in, for example, investigating the effects of strain on crystal growth.

The general theory of modelling the elastic properties of solids developed by Born & Huang (1968) has been applied to ionic systems by Catlow & Norgett (1976) and Catlow & Mackrodt (1982) and extended to the study of molecular crystals by Walmsley (1968*a*, *b*, 1987). An alternative approach is the numerical method developed by Busing & Matsui (1984), using *WMIN*, to simulate the application of external forces.

In the present paper, the numerical method of calculating the elastic constants of molecular crystals using the packing analysis program *PCK*83 (Williams, 1983), developed and applied to the model case of CO_2 in previous work (Pavlides, Pugh & Roberts 1991*a*), is extended to benzene and urea crystals. The elastic constants of CO_2 are also recalculated in an attempt to address the problem of symmetry relaxation, occurring upon distorting the lattice, in more detail. Within the numerical limitations

---

* Also at SERC Daresbury Laboratory, Warrington WA4 4AD, England.

0108-7673/91/060846-05$03.00