Acta Cryst. (1976). A**32**, 538

On the Application of Phase Relationships to Complex Structures. IX. *MULTAN* Failures

By Leslie Lessinger*

*Department of Physics, University of York, Heslington, York YO*1 5*DD, England*

(*Received* 15 *September* 1975; *accepted* 13 *January* 1976)

The failure of the multisolution direct-methods program *MULTAN* to solve six crystal structures is analysed in terms of a distinction between the general strategy and the particular tactics of solution. Consideration of the interaction between the number of large E values chosen to define the structure and the number of $\sum_2$ relationships among their phases, and the way in which *MULTAN*'s choice of a partial set of $\sum_2$ relationships limits the achievable accuracy of phase determination, leads to the conclusion that the best tactics are to use only as many E values as necessary but as many $\sum_2$ relationships as possible. Different methods of calculating E values, and the use of $\sum_1$ phase indications, are briefly discussed. The tactic of starting with more unknown phases, its effect on phase development, and the limitations set by the strategy of independent phase permutation, are examined. Finally, it is shown that there exist structures impossible to solve with *MULTAN*, probably because the tangent formula itself is inadequate, causing the phases to diverge from their true values under its operation. Directions along which to seek improved future strategies are suggested.

Introduction

The multisolution method of phase determination using convergence mapping and the tangent formula which is embodied in the computer program *MULTAN* has been very successful in solving centrosymmetric and non-centrosymmetric crystal structures containing up to 50 or 60 atoms in the asymmetric unit. Sometimes, however, *MULTAN* fails; that is, no set of phases yields an E map in which all or most of the molecule, or even a recognizable fragment, appears. It is of practical importance to go back and analyse such a case after the structure has been somehow solved and try to determine whether *MULTAN* was capable of solving that particular structure or whether other, perhaps more powerful methods were necessary. Some results from the analyses of six structures, all of which *MULTAN* initially failed to solve, are presented below.

Strategy and tactics

It is useful when discussing the successes and failures of any method of crystal structure determination to draw a distinction between the general strategy of solution, that is, the theoretical basis of the method and the algorithms used, and the particular tactics employed in a given case, that is, the actual values chosen for variable parameters and the specific path along which the solution is sought. The strategy being more or less fixed, tactics may vary according to considerations set by the particular problem to be solved, computing limitations, and previous experience or rules of thumb. The full power of a general strategy can only be exploited by the use of optimum or at least good tactics; one must also have some ideas of appropriate changes in these tactics in case of failure. A schematic listing of the steps in a structure determination using *MULTAN*, drawing this distinction explicitly, is set forth in Table 1 (Germain & Woolfson, 1968; Germain, Main & Woolfson, 1970, 1971; Declercq, Germain, Main & Woolfson, 1973; Koch, 1974; Main, Woolfson, Lessinger, Germain & Declercq, 1974).

Since *MULTAN* was designed to make structure determination as automatic as possible, the number of tactical choices is quite small. (A number of such choices are either embedded in the programs as fixed parameters or are given default values, which the user can override.) These few choices can be crucial to the success or failure of a structure determination, but rules of thumb for good choices of the parameters are not always widely known. More importantly, the reasons for particular choices and their effects on the operation of the program have not been well appreciated.

The following discussion of tactical choices will be illustrated by the examples in Tables 2–7, which for the following six structures show the conditions of either success or failure using *MULTAN*, with various figures of interest associated with each case: (1) INOS, *cis*-inositol monohydrate (Freeman, Langs, Nockolds & Oh, 1976), (2) LITH, lithocholic acid (Arora, Germain & Declerq, 1976), (3) MINA, 2-methoxyisonitrosoacetanilide (Font-Altaba, Miravitlles, Brianso, Plana & Solans, 1976), (4) RR, 3,3-dimethyl-4,5,9,10,11,12-hexacarboxymethyltetracyclo[7,2,1,$0^{2,4}$,$0^{2,8}$]dodeca-5,7,10-triene (Declercq, Germain & Henke, 1973), (5) TPH, tetraphenylhydrazine (Hoekstra, Vos, Braun & Hornstra, 1975), (6) AZET, 3-chloro-1,3,4-triphenylazetidin-2-one (Colens, Declercq, Germain, Putzeys & Van Meerssche, 1974).

---

* Present address: 14 Hamlin Street, Cambridge, Mass. 02141, U.S.A.

In order to prepare these tables, an extended and modified version of *MULTAN* was used which at each point in the procedure compares estimated or calculated quantities with their true values found by using published phases for the actual structure, possible changes of origin and/or enantiomorph from those initially defined being taken into account. In the tables which follow, phases developed by *MULTAN* are designated PHI, published phases PUB; the number of smallest E values used to calculate the ψ_0 figure of merit was 48 for RR, 50 for all other cases; the relative weights used to calculate CFOM were always taken as 1·0; the number of large peaks saved when E maps were calculated was always 2 × NAT.

Calculation of *E* values

The method used to calculate E values may itself be a decisive factor in the success or failure of a structure determination attempt by *MULTAN*, as can be seen for INOS (Table 2, case 1 *vs.* cases 5 and 6). E-set 1, used in the original unsuccessful attempt to solve the structure by direct methods, was calculated by the K-curve method (Karle, Hauptman & Christ, 1958) using the formula

$$|E_{\mathbf{h}}|^2 = K(s)\,|F_{\mathbf{h}}|^2/\varepsilon \sum f_i^2(s)$$

where $s = \sin\theta/\lambda$. For INOS, essentially the entire molecular geometry was known beforehand, and a better way of taking this into account was used to

Table 1. *Schematic listing of the steps in a structure determination using MULTAN*

Strategy	Tactics	Major considerations
(1) Compute normalized structure factors E for the entire data set with the auxiliary program *NORMAL*. Order the E values by magnitude.	Use of as much molecular geometry as is known to compute spherically averaged molecular scattering factors.	The extent to which molecular geometry is known prior to structure determination.
(2) Choose the largest E values for phase determination by *MULTAN*.	Choice of the number of large E values (NUMB).	The number of atoms in the asymmetric unit (NAT).
(3) Find all $\sum_2$ relationships among the NUMB E values. Retain the $\sum_2$ relationships with the largest values of $\kappa = 2\sigma_3\sigma_2^{-3/2}\lvert E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}\rvert$.	Choice of the number of $\sum_2$ relationships to retain (NSRT).	The total number of $\sum_2$ relationships (NSRTOT); the maximum number which can be stored (NSRMAX).
(4) Find all $\sum_1$ relationships, if any, among the NUMB E values. Accept phase indications with probability $\geq$ a limit PROB.	Choice of the $\sum_1$ probability acceptance limit PROB.	Previous experience; difficulty in solving the structure.
(5) Construct a convergence map of the NSRT $\sum_2$ relationships. This yields a starting set of phases which comprises any accepted from $\sum_1$ indications, those phases used to define the origin, to which particular values are given, and a number of unknown phases assigned multiple values which are permuted. The enantiomorph is defined by an appropriate phase restriction. The rest of the convergence map defines the order in which phase determination is attempted.	Choice of the number and type of unknown phases. The user specifies that these consist of NSPEC special phases (permutations restricted to φ_{SPEC}, $\varphi_{\mathrm{SPEC}}+180°$) + NGEN general phases (phase permutations 45°, 135°, 225°, 315°) + NANY phases of either sort.	The total number of phase sets which will need to be developed (NUMSET); the presence of weak links in the convergence map. (If special circumstances seem to warrant it, the user may specify explicitly any or all of the reflexions in the starting set rather than accepting the choices made automatically by *MULTAN*.)
(6) Determine phases using a weighted tangent formula and refine to self-consistency. Details of the slightly complicated strategy used in this subroutine (*FASTAN*) will not be discussed here.	The values of certain parameters are fixed in the program, not set by the user. These will not be discussed here.	Speed of phase development, hence the use of a weighted tangent formula; caution in the initial steps, hence the attempt to ensure that phase development follows the most probably correct path.
(7) Rank each phase set on a combined figure of merit (CFOM) based on a weighted sum of its relative absolute figure of merit, relative ψ_0 figure of merit, and relative residual.	Choice of the relative weights for ABS FOM, PSI ZERO, and RESID. Default values are 1·0, 1·0, 1·0, putting CFOM in the range 0·0–3·0.	The space group of the structure. In space groups without screw axes or glide planes, ABS FOM may be given less and PSI ZERO more weight.
(8) Using auxiliary programs, compute Fourier transforms of the most promising phase sets, search for all peaks, save the highest ones, calculate distances and angles involving these, and attempt to interpret the E maps, applying the geometrical constraints of structural organic chemistry. Strategies for the efficient interpretation of E maps, and methods for completing partial structures, will not be discussed here.	Choice of the number of E maps to calculate and examine. Choice of the number of large peaks to save. (The default action is not to print the entire E map, but rather only a projection of the peaks saved, the number of which is taken as 1·2 × NAT.)	The figures of merit of the phase sets; whether or not the combined figures of merit clearly distinguish among the phase sets. Note that various recycling procedures may sometimes succeed in extracting the correct structure even from quite unpromising fragmentary E maps.

calculate *E*-set 2, employing a modified Wilson plot based on the equation

$$|E_{\mathbf{h}}|^2 = k \exp(2Bs^2)\,|F_{\mathbf{h}}|^2/\varepsilon \sum g_m^2(s)$$

where *g* is the spherically averaged molecular scattering factor calculated from the formula (Debye, 1915)

$$g^2(s) = \sum_i \sum_j f_i(s) f_j(s) \sin(4\pi s r_{ij})/4\pi s r_{ij}\,,$$

r_{ij} = distance between atoms *i* and *j*.

The two sets of *E* values are quite different, as the following few examples show:

h	*k*	*l*	*E*−1	*E*−2
0	0	4	4·78	3·96
0	0	8	2·35	1·46
4	0	0	3·42	3·49
7	1	−16	3·26	3·55
8	1	11	2·87	3·43
1	3	13	3·96	3·04
2	3	0	3·68	3·82
7	3	−7	4·25	3·53
7	3	−3	3·00	2·66
7	3	5	3·40	2·84
5	4	−1	2·86	3·75
0	6	0	2·64	3·21
0	8	1	5·03	4·03
0	8	3	3·16	2·53
6	8	−9	2·78	3·33

In particular, reflexion 008, for which there is a strong but incorrect $\sum_1$ indication in *E*-set 1, is no longer among the 260 largest *E*-values chosen from *E*-set 2.

Table 2. INOS: $C_6H_{12}O_6.H_2O$, $P2_1/n$, $Z=8$, NAT$=12+1+12+1=26$

CASE	PARAMETERS		STARTING SET				RESULTS				MEAN PHASE ERROR	E-MAPS	
			(NO ENANTIOMORPH FIXING NEEDED)				PHASE SET CONSIDERED	FIGURES OF MERIT				CORRECT PEAKS	
			TYPE	h k l	PHI	PUB		ABS FOM	PSI ZERO	RESID		TOTAL	PEAK NUMBERS
1 (E-SET 1)	SIGMA-2		SIGMA-1	0 6 0	180, P=1.000	180	ALL 64: MAXIMUM	1.1172	-	40.15	88.6		FAILURE
	NUMB	260		0 0 8	360, P=0.999	180	MINIMUM	0.6703	-	29.58	52.1		
	NSRT	2462		4 0 0	180, P=0.992	180	CFOM 1: 2.9986	1.1166	-	29.58	84.8	-	
	NSRTOT	2462	ORIGIN DEFINING PHASES	7 3 -7	360	360	CFOM 2-9: 2.9511-	1.1159-	-	30.02-	84.5-	-	
	CONVERGE			0 8 3	180	180	2.9558	1.1172		30.10	85.2		
	PROB	0.95		1 4 -4	360	360	CFOM 10-12: 2.5698-	1.0217-	-	31.82-	82.7-	-	
	NSPEC,NGEN,NANY	0,0,6	PERMUTED PHASES	0 8 1	360,180	180	2.5744	1.0249		31.94	85.2		
	FASTAN			2 3 0	360,180	360	CFOM 13-21: 1.9980-	0.9649-	-	36.37-	52.1-	-	
	NUMSET	64		7 1 8	360,180	360	2.0289	0.9802		36.93	58.8		
				4 4-14	360,180	180	PUBLISHED PHASES	1.1809	-	18.74	0.7*		
				3 1 -4	360,180	180	PUB REFINED BY FASTAN	1.1810	-	18.80	0.0		
				5 4 -4	360,180	180							
2 (E-SET 1)	SIGMA-2		SIGMA-1	0 6 0	180, P=1.000	180	ALL 64: MAXIMUM	1.1171	-	45.71	91.4		
	NUMB	260	ORIGIN DEFINING PHASES	7 3 -7	360	360	MINIMUM	0.5335	-	24.70	56.1		
	NSRT	2462		0 8 3	180	180	CFOM 1: 2.9383	1.0811	-	24.70	66.6	12+1+	1-6,9-14,17,19,22-
	NSRTOT	2462		1 4 -4	360	360						11+1=	26,34,35,40,43,46,
	CONVERGE		PERMUTED PHASES	0 8 1	360,180	180						25	47 (Δy = -0.33)
	PROB	1.00		2 3 0	360,180	360							(PARTIAL SUCCESS)
	NSPEC,NGEN,NANY	0,0,6		4 0 0	360,180	180	CFOM 2-4: 2.7431-	1.1170-	-	30.08-	85.2-	-	
	FASTAN			7 1 8	360,180	360	2.7437	1.1171		30.10	85.2		
	NUMSET	64		3 1 -4	360,180	180	PUBLISHED PHASES	1.1809	-	18.74	0.7*		
				5 4 -4	360,180	180	PUB REFINED BY FASTAN	1.1810	-	18.80	0.0		
3 (E-SET 2)	SIGMA-2		SIGMA-1	0 6 0	180, P=1.000	180	ALL 8: MAXIMUM	1.1330	765.0	40.05	86.9		FAILURE
	NUMB	260		4 0 0	180, P=0.996	180	MINIMUM	0.6800	311.7	26.96	61.9		
	NSRT	2386	ORIGIN DEFINING PHASES	7 3 -7	360	360	CFOM 1: 2.1338	1.1330	704.4	26.96	82.4	-	
	NSRTOT	2386		0 4 11	360	360	CFOM 2: 2.0978	1.0001	494.8	29.64	83.8	-	
	CONVERGE			1 4 -4	360	360	CFOM 3: 1.3091	0.6800	311.7	36.00	86.9	-	
	PROB	0.95	PERMUTED PHASES	0 0 4	360,180	180	CFOM 4: 1.3012	0.7592	341.7	37.53	81.4	-	
	NSPEC,NGEN,NANY	0,0,3		2 3 0	360,180	360	CFOM 5: 1.2116	0.9713	765.0	32.60	61.9	-	
	FASTAN			7 4 -1	360,180	360	PUBLISHED PHASES	1.1030	209.1	19.99	0.7*		
	NUMSET	8					PUB REFINED BY FASTAN	1.1049	217.3	19.85	0.7		
4 (E-SET 2)	SIGMA-2		SIGMA-1	0 6 0	180, P=1.000	180	ALL 16: MAXIMUM	1.1334	765.0	39.73	86.6		FAILURE
	NUMB	260		4 0 0	180, P=0.996	180	MINIMUM	0.6863	303.3	26.96	61.6		
	NSRT	2386	ORIGIN DEFINING PHASES	7 3 -7	360	360	CFOM 1: 2.1305	1.1330	704.4	26.96	82.4	-	
	NSRTOT	2386		0 4 11	360	360	CFOM 2: 2.1052	1.1334	709.9	27.14	83.8	-	
	CONVERGE			1 4 -4	360	360	CFOM 3: 2.0981	1.0002	495.2	29.36	84.1	-	
	PROB	0.95	PERMUTED PHASES	0 0 4	360,180	180	CFOM 4: 2.0778	1.0001	494.6	29.63	83.8	-	
	NSPEC,NGEN,NANY	0,0,4		2 3 0	360,180	360	CFOM 5: 1.5071	0.7843	323.8	35.48	80.7	-	
	FASTAN			7 4 -1	360,180	360	PUBLISHED PHASES	1.1030	209.1	19.99	0.7*		
	NUMSET	16		3 1 -4	360,180	180	PUB REFINED BY FASTAN	1.1049	217.3	19.85	0.7		
5 (E-SET 2)	SIGMA-2		SIGMA-1	0 6 0	180, P=1.000	180	ALL 32: MAXIMUM	1.1334	765.0	41.17	88.6		
	NUMB	260		4 0 0	180, P=0.996	180	MINIMUM	0.6780	217.6	19.85	0.7		
	NSRT	2386	ORIGIN DEFINING PHASES	7 3 -7	360	360	CFOM 1: 2.9371	1.1048	217.6	19.85	0.7	26	1-22,24,28,30,46
	NSRTOT	2386		0 4 11	360	360	CFOM 2: 2.9223	1.1047	222.9	19.95	1.4	26	1-18,20-23,25,31,
	CONVERGE			1 4 -4	360	360							32,44
	PROB	0.95	PERMUTED PHASES	0 0 4	360,180	180	CFOM 3: 2.2007	1.0335	357.5	26.77	72.0	-	
	NSPEC,NGEN,NANY	0,0,5		2 3 0	360,180	360	CFOM 4: 2.1931	1.0335	360.8	26.80	72.0	-	
	FASTAN			7 4 -1	360,180	360	PUBLISHED PHASES	1.1030	209.1	19.99	0.7*		
	NUMSET	32		3 1 -4	360,180	180	PUB REFINED BY FASTAN	1.1049	217.3	19.85	0.7		
				7 3 1	360,180	180							
6 (E-SET 2)	SIGMA-2		SIGMA-1	0 6 0	180, P=1.000	180	ALL 64: MAXIMUM	1.1334	765.0	45.22	88.6		
	NUMB	260		4 0 0	180, P=0.996	180	MINIMUM	0.5213	217.6	19.85	0.7		
	NSRT	2386	ORIGIN DEFINING PHASES	7 3 -7	360	360	CFOM 1,2: 2.9532	1.1048	217.6	19.85	0.7	26	1-22,24,28,30,46
	NSRTOT	2386		0 4 11	360	360	CFOM 3,4: 2.9392	1.1047	222.9	19.95	1.4	26	1-18,20-23,25,31,
	CONVERGE			1 4 -4	360	360							32,44
	PROB	0.95	PERMUTED PHASES	0 0 4	360,180	180	CFOM 5-8: 2.3013-	1.0334-	357.5-	26.74-	72.0-	-	
	NSPEC,NGEN,NANY	0,0,6		2 3 0	360,180	360	2.3097	1.0337	360.8	26.80	72.0		
	FASTAN			7 4 -1	360,180	360	PUBLISHED PHASES	1.1030	209.1	19.99	0.7*		
	NUMSET	64		7 3 5	360,180	360	PUB REFINED BY FASTAN	1.1049	217.3	19.85	0.7		
				3 1 -4	360,180	180							
				7 3 1	360,180	180							

Choice of *E* and definition of the problem

Assuming that we have calculated *E* values as best we can, *MULTAN* now makes the simplest choice among them, selecting the NUMB largest. None of the other structure factors is considered at all in the phase determination, and it is only afterwards that ψ_0, which uses the smallest *E* values, is calculated as a figure of merit. Thus *MULTAN* is in effect determining phases for a 'structure' which is *defined* as the Fourier transform of the NUMB chosen *E* values taken with the phases of the true structure, all other *E* values being taken as having zero magnitude. Since the user *chooses* this structure, he must know what properties it should have and how to ensure that it does have them.

The crucial property is that a 1:1 correspondence can easily be made between the true structure and the defined structure; that is, the true atomic positions are recognizable among the highest peaks of the defined structure. Awareness of the necessity for this correspondence is usually expressed by asking how *many* *E* values are needed to make a good Fourier map of the structure. A common rule of thumb is to use at least 10 *E* values per atom to be found, which, if the phases are good, will certainly ensure that the true atomic positions are well defined in the *E* map.

It is important, though, to ask as well how *few* *E* values will suffice to make a recognizable Fourier map of the structure. The number of *E* values chosen may be a critical number and not just a matter of convenience when using *MULTAN*, and as will become clear from the discussion below, one wishes to take only as many *E* values as necessary. From the figures for LITH (Table 3, case 1), MINA (Table 4, case 1), TPH (Table 6, case 1), and, among other recent examples, 4-octadecynoic acid ($C_{18}H_{32}O_2$), crocetindi-

Table 3. LITH: $C_{24}H_{40}O_3$, $P2_12_12_1$, $Z=4$, NAT=27

CASE	PARAMETERS		STARTING SET				RESULTS				MEAN PHASE ERROR	E-MAPS	
			ENANTIOMORPH FIXED BY (E)				PHASE SET CONSIDERED	FIGURES OF MERIT				CORRECT PEAKS	
			TYPE	h k l	PHI	PUB		ABS FOM	PSI ZERO	RESID		TOTAL	PEAK NUMBERS
1	SIGMA-2		SIGMA-1	NONE	WITH P ≥ 0.95	-	ALL 32: MAXIMUM	1.0915	334.1	38.59	87.2		
	NUMB	199	ORIGIN	2 0 11	90	90	MINIMUM	0.6558	161.8	17.35	9.5		
	NSRT	1975	DEFINING	0 11 8	90	90	CFOM 1: 2.9958	1.0897	161.8	17.35	9.5	24	1-7,9,11-14,16,18,20,21,23,25,28,31,32,34,39,40
	NSRTOT	1975	PHASES	3 5 0	90	90							
	CONVERGE			2 3 17	45,135,225,315	83							
	PROB	0.95	PERMUTED	4 6 12	45,135,225,315	100	CFOM 2: 2.9842	1.0915	162.3	17.61	9.6	25	1-8,11-15,18-20,23,24,30,31,35-37,40,44
	NSPEC,NGEN,NANY	0,0,3	PHASES	1 1 1	45,135 (E)	95							
	FASTAN												
	NUMSET	32					CFOM 3: 1.7142	0.8919	202.5	29.91	33.4	23	1-4,6-8,10-13,16,17,21,22,24,25,29,34,36,37,41,42,44
							CFOM 4: 1.6026	0.9888	269.3	28.76	86.2	-	
							PUBLISHED PHASES	1.0363	161.7	16.42	6.9*		
							PUB REFINED BY FASTAN	1.0973	162.6	17.21	9.4		
2	SIGMA-2		SIGMA-1	NONE	WITH P ≥ 0.95	-	ALL 32: MAXIMUM	0.9604	406.6	44.80	88.1		FAILURE
	NUMB	270	ORIGIN	2 0 11	90	90	MINIMUM	0.6990	212.0	37.85	62.5		
	NSRT	1620	DEFINING	0 11 8	90	90	CFOM 1: 2.3613	0.9338	306.1	38.22	83.9	-	
	NSRTOT	5062	PHASES	3 5 0	90	90	CFOM 2: 2.2748	0.9030	310.3	37.85	88.1	-	
	CONVERGE			1 1 12	45,315 (E)	338	CFOM 3: 2.0046	0.8793	277.0	40.29	76.1	-	
	PROB	0.95	PERMUTED	2 3 17	45,135,225,315	83	CFOM 4: 1.8552	0.8793	281.5	41.17	77.6	-	
	NSPEC,NGEN,NANY	0,0,3	PHASES	1 1 1	45,135,225,315	95	CFOM 5: 1.8261	0.8745	289.2	40.96	73.7	-	
	FASTAN						PUBLISHED PHASES	1.0807	164.8	32.35	10.5*		
	NUMSET	32					PUB REFINED BY FASTAN	1.1330	172.2	32.28	11.7		
3	SIGMA-2		SIGMA-1	NONE	WITH P ≥ 0.95	-	ALL 128: MAXIMUM	1.1330	443.6	45.31	89.0		
	NUMB	270	ORIGIN	2 0 11	90	90	MINIMUM	0.6205	170.8	32.40	11.7		
	NSRT	1620	DEFINING	0 11 8	90	90	CFOM 1: 2.9938	1.1330	172.5	32.40	11.7	27	1-12,14-18,20-24,28,30,35,38,40
	NSRTOT	5062	PHASES	3 5 0	90	90							
	CONVERGE			1 1 12	45,315 (E)	338	CFOM 2: 2.9467	1.1291	170.8	32.99	12.6	27	1-9,11-19,21,22,24,26,32,33.37,40,41
	PROB	0.95	PERMUTED	2 3 17	45,135,225,315	83							
	NSPEC,NGEN,NANY	0,0,4	PHASES	1 1 1	45,135,225,315	95							
	FASTAN			1 2 18	45,135,225,315	94	CFOM 3: 1.7512	0.9097	272.0	38.11	87.8	-	
	NUMSET	128					PUBLISHED PHASES	1.0807	164.8	32.35	10.5*		
							PUB REFINED BY FASTAN	1.1330	172.2	32.28	11.7		
4	SIGMA-2		SIGMA-1	NONE	WITH P ≥ 0.95	-	ALL 32: MAXIMUM	1.0071	616.2	41.58	88.4		FAILURE
	NUMB	270	ORIGIN	3 5 0	90	90	MINIMUM	0.7462	355.7	35.45	50.4		
	NSRT	4000	DEFINING	0 2 23	180	180	CFOM 1: 2.2710	0.8842	384.7	36.35	65.3	-	
	NSRTOT	5062	PHASES	0 3 17	90	90	CFOM 2: 2.1204	0.8518	404.6	36.04	80.7	-	
	CONVERGE			1 1 12	45,315 (E)	338	CFOM 3: 2.0815	0.8517	391.6	36.58	79.3	-	
	PROB	0.95	PERMUTED	2 4 1	45,135,225,315	208	CFOM 4: 2.0310	0.9505	498.3	36.70	64.2	-	
	NSPEC,NGEN,NANY	0,0,3	PHASES	2 4 18	45,135,225,315	240	CFOM 5: 1.8681	0.7462	390.0	35.45	86.5	-	
	FASTAN						PUBLISHED PHASES	1.0312	200.0	19.32	6.2*		
	NUMSET	32					PUB REFINED BY FASTAN	1.0737	201.0	20.22	8.3		
5	SIGMA-2		SIGMA-1	NONE	WITH P ≥ 0.95	-	ALL 128: MAXIMUM	1.0798	636.2	41.72	88.3		
	NUMB	270	ORIGIN	3 5 0	90	90	MINIMUM	0.6434	201.1	20.29	8.4		
	NSRT	4000	DEFINING	0 2 23	180	180	CFOM 1: 2.9857	1.0739	201.5	20.29	8.4	27	1-19,21-23,35,36,42,46,50
	NSRTOT	5062	PHASES	0 3 17	90	90							
	CONVERGE			1 1 12	45,315 (E)	338	CFOM 2: 2.9845	1.0731	201.1	20.30	8.4	27	1-18,20-22,24,35,36,42,45,51
	PROB	0.95	PERMUTED	2 4 1	45,135,225,315	208							
	NSPEC,NGEN,NANY	0,0,4	PHASES	1 12 4	45,135,225,315	210	CFOM 3: 2.9831	1.0743	201.7	20.35	8.6	27	1-18,20-23,32,33,40,43,45
	FASTAN			2 4 18	45,135,225,315	240							
	NUMSET	128					CFOM 4: 2.9829	1.0738	201.6	20.33	8.4	27	1-19,21-23,33,36,42,45,50
							CFOM 5: 1.6098	0.9388	399.0	33.41	86.4	-	
							PUBLISHED PHASES	1.0312	200.0	19.32	6.2*		
							PUB REFINED BY FASTAN	1.0737	201.0	20.22	8.3		

aldehyde ($C_{20}H_{24}O_2$), δ-pyrazinecarboxamide ($C_5H_5N_3O$) (Mo, Hjortås & Svinning, 1973), and 1,2:3,4:5,6-tris-(*o*,*o*′-biphenylene)borazine ($C_{36}H_{24}B_3N_3$) (Roberts, Brauer, Tsay & Krüger, 1974), it is clear that the use of five to seven *E* values per atom to be determined may be quite sufficient, if the phases are close to their true values, to give a good, readily interpretable Fourier map, and that ten *E* values per atom should perhaps better be thought of as an upper limit. [Note that when the correspondence between the defined and the true structure, judged by the peak numbers of correct peaks in *E* maps made with published phases, is somewhat poor, as for RR (Table 5, case 1) and AZET (Table 7, cases 1 and 2), there is greater difficulty, as might be expected, in developing good phases].

Choice of $\sum_2$ and limitations on the solution

The second property required of the defined structure is that the assumptions made in deriving the direct-methods formulae and phase relationships used by *MULTAN* hold sufficiently well for these to be applied to the defined structure. Essentially we require that Sayre's equation, and the related tangent formula, apply. When the number of *E* values per atom in the asymmetric unit is larger than about 5, Sayre's equation or at least its angular part does in fact hold to a high degree of approximation, *when all the terms are included in the summation.* This can be seen by applying the tangent formula once to *calculate, simultaneously*, the phases of the NUMB largest *E* values from their *published* phases, including all terms:

Table 4. MINA: $C_9H_{10}N_2O_3$, $P2_12_12_1$, $Z=8$, NAT$=14+14=28$

CASE	PARAMETERS		STARTING SET ENANTIOMORPH FIXED BY (E) TYPE	h k l	PHI	PUB
1	SIGMA-2		SIGMA-1	NONE	WITH P $\geqslant$ 0.95	-
	NUMB	182	ORIGIN DEFINING PHASES	17 3 0	270	270
	NSRT	1505		0 14 3	360	360
	NSRTOT	1505		7 18 0	90	90
	CONVERGE		PERMUTED PHASES	13 0 2	360,180	180
	PROB	0.95		12 3 2	45,135,225,315	341
	NSPEC,NGEN,NANY	0,0,3		12 1 2	45,135 (E)	29
	FASTAN					
	NUMSET	16				
2	SIGMA-2		SIGMA-1	NONE	WITH P $\geqslant$ 0.95	-
	NUMB	280	ORIGIN DEFINING PHASES	17 3 0	270	270
	NSRT	1540		7 18 0	90	90
	NSRTOT	5220		0 5 1	90	90
	CONVERGE		PERMUTED PHASES	13 0 2	180 (E)	180
	PROB	0.95		8 18 0	360,180	360
	NSPEC,NGEN,NANY	0,0,3		12 8 4	45,135,225,315	76
	FASTAN					
	NUMSET	8				
3	SIGMA-2		SIGMA-1	NONE	WITH P $\geqslant$ 0.95	-
	NUMB	280	ORIGIN DEFINING PHASES	17 3 0	270	270
	NSRT	1540		7 18 0	90	90
	NSRTOT	5220		0 5 1	90	90
	CONVERGE		PERMUTED PHASES	13 0 2	360,180	180
	PROB	0.95		8 18 0	360,180	360
	NSPEC,NGEN,NANY	0,0,4		12 8 4	45,135,225,315	76
	FASTAN			1 4 1	225,315 (E)	337
	NUMSET	32				
4	SIGMA-2		SIGMA-1	NONE	WITH P $\geqslant$ 0.95	-
	NUMB	280	ORIGIN DEFINING PHASES	17 3 0	270	270
	NSRT	2000		0 14 3	360	360
	NSRTOT	5220		0 5 1	90	90
	CONVERGE		PERMUTED PHASES	13 0 2	360,180	180
	PROB	0.95		12 3 2	45,315 (E)	341
	NSPEC,NGEN,NANY	0,0,3		12 1 2	45,135,225,315	29
	FASTAN					
	NUMSET	16				
5	SIGMA-2		SIGMA-1	NONE	WITH P $\geqslant$ 0.95	-
	NUMB	280	ORIGIN DEFINING PHASES	17 3 0	270	270
	NSRT	4000		0 14 3	360	360
	NSRTOT	5220		12 17 0	180	180
	CONVERGE		PERMUTED PHASES	13 0 2	360,180	180
	PROB	0.95		12 3 2	225,315 (E)	341
	NSPEC,NGEN,NANY	0,0,3		12 1 2	45,135,225,315	29
	FASTAN					
	NUMSET	16				

CASE	RESULTS PHASE SET CONSIDERED	FIGURES OF MERIT ABS FOM	PSI ZERO	RESID	MEAN PHASE ERROR	E-MAPS CORRECT PEAKS TOTAL	PEAK NUMBERS
1	ALL 16: MAXIMUM	1.0867	295.1	34.90	86.0		
	MINIMUM	0.7235	162.1	15.26	9.4		
	CFOM 1: 2.9991	1.0867	162.2	15.26	9.4	13+12 =25	1-6,8-17,21,23,26,28-32,36
	CFOM 2: 2.6116	1.0383	162.1	20.27	13.5	14+11	1-12,14,15,18,19,21,23,29,31,35,37,40,44,45
	CFOM 3: 1.2606	0.7717	168.9	31.38	53.5		
	PUBLISHED PHASES	1.0129	157.4	15.79	7.9*		
	PUB REFINED BY FASTAN	1.0860	162.1	15.62	9.3		
2	ALL 8: MAXIMUM	0.8974	350.2	46.44	87.3		
	MINIMUM	0.5747	183.6	40.91	60.0		
	CFOM 1: 2.0861	0.8974	335.8	40.91	82.2	-	FAILURE
	CFOM 2: 1.8000	0.8567	331.1	41.95	79.5	-	
	CFOM 3: 1.6491	0.8568	350.2	42.15	82.6	-	
	CFOM 4: 1.5900	0.6644	219.9	43.50	81.2	-	
	CFOM 5: 1.5283	0.8402	339.2	42.90	79.7	-	
	PUBLISHED PHASES	1.0933	214.8	34.41	11.5*		
	PUB REFINED BY FASTAN	1.1607	208.7	33.40	11.7		
3	ALL 32: MAXIMUM	1.1612	387.2	47.58	86.8		
	MINIMUM	0.6240	169.2	33.34	11.7		
	CFOM 1: 2.8230	1.1612	207.8	33.34	11.8	14+13 =27	1-15,17-24,26,34,46,56
	CFOM 2: 2.8187	1.1612	208.3	33.36	11.7	14+12 =26	1-16,18-24,26,33,46
	CFOM 3: 2.8142	1.1606	208.4	33.41	11.8	14+13 =27	1-16,18-25,34,44,52
	CFOM 4: 2.8068	1.1602	208.5	33.49	11.8	14+12 =26	1-16,18-24,26,34,46
	CFOM 5: 2.7925	1.1594	210.1	33.58	11.8	14+13 =27	1-16,18-24,26,34,42,54
	CFOM 6-8: 1.3666-1.4226	0.6795-0.7526	192.0-207.1	41.54-43.20	84.1-85.0	-	
	CFOM 9: 1.3456	0.8753	280.8	42.03	38.6	13+13 =26	1-5,7-10,12-14,17,19,25,28,31,34,36-39,43,44,47,55
	CFOM 10: 1.3177	0.8756	270.1	43.14	46.2	11+2 =13	1-4,7,8,16,22,25,33,36,37,42
	CFOM 11: 1.2786	0.7284	203.8	44.12	79.8	-	
	PUBLISHED PHASES	1.0933	214.8	34.41	11.5*		
	PUB REFINED BY FASTAN	1.1608	208.7	33.37	11.7		
4	ALL 16: MAXIMUM	1.1485	499.8	47.68	84.9		
	MINIMUM	0.6442	216.6	29.60	11.2		
	CFOM 1: 2.9883	1.1485	219.9	29.60	11.2	14+14 =28	1-14,16-24,26,27,29,32,37
	CFOM 2: 1.6578	0.8229	216.6	42.19	42.4	12+7 =19	1-3,6-9,12,13,18,19,21,22,25,26,28,29,39,42
	CFOM 3: 1.3550	0.6714	222.8	41.83	84.9	-	
	PUBLISHED PHASES	1.0773	223.7	31.10	10.4*		
	PUB REFINED BY FASTAN	1.1487	219.6	29.58	11.1		
5	ALL 16: MAXIMUM	0.9624	519.2	40.60	87.1		
	MINIMUM	0.7937	278.7	34.38	36.1		
	CFOM 1: 2.1656	0.8216	278.7	34.38	36.1	13+12 =25	1-12,17-19,21,23,25-27,30,34,36,37,56
	CFOM 2: 1.9847	0.8174	308.3	34.58	36.8	13+11 =24	1-10,13-15,20,22-24,27,29,34,35,42,46,53
	CFOM 3: 1.5994	0.9534	445.2	38.45	77.6	-	
	PUBLISHED PHASES	1.0523	228.0	21.21	7.8*		
	PUB REFINED BY FASTAN	1.1180	236.5	21.73	9.3		

$$\tan\varphi_{\mathbf{h}}(\mathrm{CALC})=\frac{\sum_{\mathbf{k}}\kappa_{\mathbf{hk}}\sin[\varphi_{\mathbf{k}}(\mathrm{PUB})+\varphi_{\mathbf{h}-\mathbf{k}}(\mathrm{PUB})]}{\sum_{\mathbf{k}}\kappa_{\mathbf{hk}}\cos[\varphi_{\mathbf{k}}(\mathrm{PUB})+\varphi_{\mathbf{h}-\mathbf{k}}(\mathrm{PUB})]},$$

(NUMB such equations).

Note that this is not a stepwise or iterative procedure, so that each calculated phase is independent of all the others. The phases returned differ only slightly, on average, from the input published phases. The mean absolute value of the difference φ(PUB)-

Table 5. RR: $C_{26}H_{28}O_{12}$, $Pn2_1a$, $Z=4$, NAT=38

CASE	PARAMETERS		STARTING SET						RESULTS				MEAN PHASE ERROR	E-MAPS	
			ENANTIOMORPH FIXED BY (E)						PHASE SET CONSIDERED	FIGURES OF MERIT				CORRECT PEAKS	
			TYPE	h	k	l	PHI	PUB		ABS FOM	PSI ZERO	RESID		TOTAL	PEAK NUMBERS
1	SIGMA-2		SIGMA-1	6	0	0	180, P=0.997	180	ALL 32: MAXIMUM	1.2500	383.6	35.46	73.7		FAILURE
	NUMB	221		12	0	0	180, P=0.983	180	MINIMUM	1.1438	327.1	32.48	61.8		
	NSRT	1736		0	0	8	360, P=0.951	360	CFOM 1: 2.2133	1.2317	361.8	32.48	62.0	-	
	NSRTOT	1736	ORIGIN	6	8	1	275	275	CFOM 2-22,24-31:	1.2172-	370.0-	33.00-	61.8-	-	
	CONVERGE		DEFINING	0	9	7	45 (E)	13	1.2016-2.0220	1.2500	380.1	34.91	69.9		
	PROB	0.95	PHASES	1	3	3	45,315	33	CFOM 23: 1.5322	1.1438	327.1	33.88	73.7	-	
	NSPEC,NGEN,NANY	0,0,2	PERMUTED	2	5	3	45,135,225,315	27	CFOM 32: 0.5723	1.2046	383.6	35.46	65.7	-	
	FASTAN		PHASES	3	7	5	45,135,225,315	212	PUBLISHED PHASES	0.9495	151.5	23.21	13.3•	33	1-18,22-25,28-31, 33,39,57,62,70,71, 75
	NUMSET	32							PUB REFINED BY FASTAN	1.0816	172.9	21.48	16.7		
2	SIGMA-2		SIGMA-1	6	0	0	180, P=0.997	180	ALL 128: MAXIMUM	1.2515	386.2	35.46	77.0		FAILURE
	NUMB	221		12	0	0	180, P=0.983	180	MINIMUM	0.8195	189.5	25.84	56.5		
	NSRT	1736		0	0	8	360, P=0.951	360	CFOM 1: 1.9768	0.8196	194.1	25.84	70.7	-	
	NSRTOT	1736	ORIGIN	6	8	1	275	275	CFOM 2: 1.9630	0.8711	189.5	27.35	68.2	-	
	CONVERGE		DEFINING	0	9	7	45 (E)	13	CFOM 3: 1.8275	0.9533	220.2	28.98	58.3	-	
	PROB	0.95	PHASES	1	3	3	45,315	33	CFOM 4: 1.6872	1.1547	322.9	29.79	74.4	-	
	NSPEC,NGEN,NANY	0,0,3		2	5	3	45,135,225,315	27	CFOM 5: 1.6547	0.8924	229.9	28.81	67.4	-	
	FASTAN		PERMUTED	3	7	5	45,135,225,315	212	PUBLISHED PHASES	0.9495	151.5	23.21	13.3•	33	AS IN CASE 1
	NUMSET	128	PHASES	4	9	4	45,135,225,315	353	PUB REFINED BY FASTAN	1.0812	172.9	21.53	16.8		
3	SIGMA-2		SIGMA-1	0	0	4	360, P=0.999	360	ALL 32: MAXIMUM	1.2584	523.5	42.76	79.6		FAILURE
	NUMB	402		6	0	0	180, P=0.998	180	MINIMUM	0.8391	203.3	38.65	64.3		
	NSRT	2000		12	0	0	180, P=0.998	180	CFOM 1: 2.0609	1.1351	409.9	38.65	70.5	-	
	NSRTOT	9372		0	0	8	360, P=0.974	360	CFOM 2: 1.7739	1.1148	416.0	39.56	75.3	-	
	CONVERGE		ORIGIN	2	5	3	27	27	CFOM 3: 1.7707	1.2428	493.7	39.83	71.5	-	
	PROB	0.95	DEFINING	7	11	8	45 (E)	15	CFOM 4: 1.5979	1.2435	498.0	40.49	67.9	-	
	NSPEC,NGEN,NANY	0,0,2	PHASES	5	17	7	45,315	302	CFOM 5: 1.5715	1.2534	501.0	40.65	66.9	-	
	FASTAN		PERMUTED	4	9	4	45,135,225,315	353	PUBLISHED PHASES	0.9602	150.7	38.17	22.8•		
	NUMSET	32	PHASES	8	10	2	45,135,225,315	300	PUB REFINED BY FASTAN	1.0820	184.5	35.74	27.1		
4	SIGMA-2		SIGMA-1	0	0	4	360, P=0.999	360	ALL 128: MAXIMUM	1.2661	521.4	43.76	80.7		FAILURE
	NUMB	402		6	0	0	180, P=0.998	180	MINIMUM	0.7942	191.7	38.98	57.3		
	NSRT	2000		12	0	0	180, P=0.998	180	CFOM 1: 2.0163	1.1088	406.2	38.98	73.5	-	
	NSRTOT	9372		0	0	8	360, P=0.974	360	CFOM 2: 1.9797	1.1104	415.5	39.03	75.7	-	
	CONVERGE		ORIGIN	2	5	3	27	27	CFOM 3: 1.9254	1.2536	497.5	39.56	68.2	-	
	PROB	0.95	DEFINING	7	11	8	45 (E)	15	CFOM 4: 1.9196	1.1119	433.6	39.07	80.7	-	
	NSPEC,NGEN,NANY	0,0,3	PHASES	5	17	7	45,315	302	CFOM 5: 1.8886	1.2661	507.7	39.71	66.1	-	
	FASTAN			4	9	4	45,135,225,315	353	PUBLISHED PHASES	0.9602	150.7	38.17	22.8•		
	NUMSET	128	PERMUTED	8	10	2	45,135,225,315	300	PUB REFINED BY FASTAN	1.0816	183.2	35.68	27.1		
			PHASES	10	11	5	45,135,225,315	301							
5	SIGMA-2		SIGMA-1	0	0	4	360, P=0.999	360	ALL 32: MAXIMUM	1.2398	815.9	42.59	73.0		
	NUMB	402		6	0	0	180, P=0.998	180	MINIMUM	1.0022	327.3	31.95	26.8		
	NSRT	4000		12	0	0	180, P=0.998	180	CFOM 1: 2.1526	1.0385	327.3	31.95	26.8	35	1-12,14-27,31,33, 36,41,45-48,69
	NSRTOT	9372		0	0	8	360, P=0.974	360							
	CONVERGE		ORIGIN	1	3	3	33	33	CFOM 2: 1.8248	1.0082	406.7	32.35	36.7	32	1-9,11-13,15-18, 20-23,28,29,31-33, 36,38,39,41,59,70, 74
	PROB	0.95	DEFINING	6	8	1	315 (E)	275							
	NSPEC,NGEN,NANY	0,0,2	PHASES	2	9	7	45,315	67							
	FASTAN		PERMUTED	2	10	2	45,135,225,315	95							
	NUMSET	32	PHASES	4	12	7	45,135,225,315	193	CFOM 3: 1.5159	1.0060	415.7	35.35	73.0	-	
									PUBLISHED PHASES	0.9105	216.7	32.21	16.5•		
									PUB REFINED BY FASTAN	1.0458	301.8	31.44	22.1		
6	SIGMA-2		SIGMA-1	0	0	4	360, P=0.999	360	ALL 128: MAXIMUM	1.2384	815.6	42.70	73.0		FAILURE
	NUMB	402		6	0	0	180, P=0.998	180	MINIMUM	0.8524	459.4	36.18	61.7		
	NSRT	4000		12	0	0	180, P=0.998	180	CFOM 1: 2.0000	0.8524	459.4	36.18	66.9	-	
	NSRTOT	9372		0	0	8	360, P=0.974	360	CFOM 2: 1.6281	0.9386	568.1	38.07	71.5	-	
	CONVERGE		ORIGIN	1	3	3	33	33	CFOM 3: 1.6125	0.9073	536.3	38.22	71.2	-	
	PROB	0.95	DEFINING	6	8	1	315 (E)	275	CFOM 4-127: 1.1589-	1.1994-	781.6-	39.27-	61.7-	-	
	NSPEC,NGEN,NANY	0,0,3	PHASES	2	9	7	45,315	67	1.5858	1.2384	815.6	41.49	73.0		
	FASTAN			2	10	2	45,135,225,315	95	CFOM 128: 0.5623	0.9067	665.3	42.70	71.0	-	
	NUMSET	128	PERMUTED	4	9	4	45,135,225,315	353	PUBLISHED PHASES	0.9105	216.7	32.21	16.5•		
			PHASES	4	12	7	45,135,225,315	193	PUB REFINED BY FASTAN	1.0452	300.0	31.41	21.9		
7	SIGMA-2		SIGMA-1	0	0	4	360, P=0.999	360	ALL 32: MAXIMUM	1.2163	835.7	46.36	75.7		FAILURE
	NUMB	402		6	0	0	180, P=0.998	180	MINIMUM	0.8517	585.3	40.01	57.3		
	NSRT	8000		12	0	0	180, P=0.998	180	CFOM 1: 2.0453	0.8682	585.3	40.01	69.7	-	
	NSRTOT	9372		0	0	8	360, P=0.974	360	CFOM 2-5: 1.7571-	1.2144-	825.1-	40.86-	67.6-	-	
	CONVERGE		ORIGIN	6	8	1	275	275	1.9026	1.2163	835.7	41.54	73.1		
	PROB	0.95	DEFINING	1	3	3	45 (E)	33	CFOM 6: 1.7107	0.8517	590.5	41.71	73.1	-	
	NSPEC,NGEN,NANY	0,0,2	PHASES	2	9	7	45,315	67	CFOM 7: 1.5620	0.8712	630.1	41.99	57.3	-	
	FASTAN		PERMUTED	4	9	4	45,135,225,315	353	PUBLISHED PHASES	0.8634	223.7	24.16	11.6•		
	NUMSET	32	PHASES	8	10	2	45,135,225,315	300	PUB REFINED BY FASTAN	0.9467	284.6	24.17	15.8		
8	SIGMA-2		SIGMA-1	0	0	4	360, P=0.999	360	ALL 128: MAXIMUM	1.2230	842.9	46.43	77.9		FAILURE
	NUMB	402		6	0	0	180, P=0.998	180	MINIMUM	0.8110	523.1	36.20	59.4		
	NSRT	8000		12	0	0	180, P=0.998	180	CFOM 1: 2.0018	0.9698	645.8	36.20	73.0	-	
	NSRTOT	9372		0	0	8	360, P=0.974	360	CFOM 2: 1.9409	0.9644	646.2	36.68	73.4	-	
	CONVERGE		ORIGIN	6	8	1	275	275	CFOM 3: 1.8194	0.8343	523.1	38.63	74.7	-	
	PROB	0.95	DEFINING	1	3	3	45 (E)	33	CFOM 4: 1.7977	0.8313	523.7	38.75	76.5	-	
	NSPEC,NGEN,NANY	0,0,3	PHASES	2	9	7	45,315	67	CFOM 5: 1.6385	1.2222	828.4	40.34	68.2	-	
	FASTAN			4	9	4	45,135,225,315	353	PUBLISHED PHASES	0.8634	223.7	24.16	11.6•		
	NUMSET	128	PERMUTED	8	10	2	45,135,225,315	300	PUB REFINED BY FASTAN	0.9466	284.1	24.15	15.8		
			PHASES	10	15	5	45,135,225,315	355							

φ(CALC) (the mean tangent formula phase shift) is given in Tables 2–7 for each value of NUMB and NSRT as the 'mean phase error' for the published phases, and marked by an asterisk. To provide true comparison with the results from *MULTAN*, when a phase is restricted to only two possible values, φ(PUB)-φ(CALC) is taken as zero [or 180°, if φ(PUB)-φ(CALC) exceeds 90°].

When not all the terms are included, the tangent formula still gives the most *probable* value for a phase, estimated from the terms available, but as the number of terms omitted from the summation increases, the chance of the most probable value differing widely from the true value also increases. Thus the choice of NUMB must take into account another consideration. As NUMB increases for a given structure, the total number of $\sum_2$ relationships increases very rapidly, approximately as (NUMB)^3. The amount of space available to store these in the computer is limited, storage space for the $\sum_2$ relationships forming the

Table 6. TPH: $C_{24}H_{20}N_2$, $C222_1$, $Z=12$, NAT=26+13=39

CASE	PARAMETERS		STARTING SET				RESULTS				MEAN PHASE ERROR	E-MAPS	
			ENANTIOMORPH FIXED BY (E)				PHASE SET CONSIDERED	FIGURES OF MERIT				CORRECT PEAKS	
			TYPE	h k l	PHI	PUB		ABS FOM	PSI ZERO	RESID		TOTAL	PEAK NUMBERS
1	SIGMA-2		SIGMA-1	2 6 0	180, P=0.977	180	ALL 32: MAXIMUM	1.1123	352.4	36.26	86.4		
	NUMB	243	ORIGIN	3 45 0	360	360	MINIMUM	0.7283	151.1	18.88	9.2		
	NSRT	1999	DEFINING	1 3 9	315 (E)	311	CFOM 1: 2.7745	1.0257	151.1	18.88	9.2	25+13 =38	1-13,15-18,20-25; 27,29-33,39-41,46, 47,54,59,60,62
	NSRTOT	1999	PHASES										
	CONVERGE			12 26 6	45,135,225,315	222							
	PROB	0.95	PERMUTED	3 15 2	45,135,225,315	135	CFOM 2: 1.6568	0.9031	206.8	27.95	62.9	-	
	NSPEC,NGEN,NANY	0,0,3	PHASES	13 3 0	360,180	180	CFOM 3: 1.6221	0.8909	204.9	28.16	66.5	-	
	FASTAN						PUBLISHED PHASES	0.9662	138.6	19.96	8.4*		
	NUMSET	32					PUB REFINED BY FASTAN	1.0267	151.7	19.12	9.3		
2	SIGMA-2		SIGMA-1	2 6 0	180, P=0.977	180	ALL 16: MAXIMUM	1.0823	452.9	43.70	87.2		
	NUMB	390	ORIGIN	3 45 0	360	360	MINIMUM	0.7655	187.0	37.31	70.8		
	NSRT	2000	DEFINING	1 3 9	315 (E)	311	CFOM 1: 2.0671	0.8981	231.6	38.48	70.8	-	FAILURE
	NSRTOT	7558	PHASES				CFOM 2: 2.0356	1.0823	443.5	37.31	85.3	-	
	CONVERGE			0 30 2	360,180	180	CFOM 3: 1.9019	0.9430	268.5	39.55	86.9	-	
	PROB	0.95	PERMUTED	12 26 6	45,135,225,315	222	CFOM 4: 1.8820	0.8627	196.6	39.79	82.8	-	
	NSPEC,NGEN,NANY	0,0,3	PHASES	12 36 0	360.180	360	CFOM 5: 1.8548	1.0685	452.9	37.96	84.6	-	
	FASTAN						PUBLISHED PHASES	0.9453	172.3	36.31	16.5*		
	NUMSET	16					PUB REFINED BY FASTAN	1.0039	179.9	35.22	20.5		
3	SIGMA-2		SIGMA-1	2 6 0	180, P=0.977	180	ALL 64: MAXIMUM	1.1610	520.2	45.33	89.7		
	NUMB	390	ORIGIN	3 45 0	360	360	MINIMUM	0.7548	171.8	37.07	63.5		
	NSRT	2000	DEFINING	1 3 9	315 (E)	311	CFOM 1: 2.0822	0.9054	246.6	37.68	71.2	-	FAILURE
	NSRTOT	7558	PHASES				CFOM 2: 2.0407	1.0877	443.1	37.07	84.9	-	
	CONVERGE			0 30 2	360,180	180	CFOM 3: 2.0078	0.9418	265.9	38.57	86.7	-	
	PROB	0.95	PERMUTED	12 26 6	45,135,225,315	222	CFOM 4: 1.9683	0.9342	267.2	38.71	87.1	-	
	NSPEC,NGEN,NANY	2,0,2	PHASES	12 36 0	360,180	360	CFOM 5: 1.9611	1.0808	448.3	37.46	84.8	-	
	FASTAN			4 2 5	45,135,225,315	115	PUBLISHED PHASES	0.9453	172.3	36.31	16.5*		
	NUMSET	64					PUB REFINED BY FASTAN	1.0044	179.9	35.19	20.5		
4	SIGMA-2		SIGMA-1	2 6 0	180, P=0.977	180	ALL 32: MAXIMUM	1.2568	787.3	40.55	89.7		
	NUMB	390	ORIGIN	3 45 0	360	360	MINIMUM	0.7681	309.2	34.73	71.8		
	NSRT	4000	DEFINING	1 3 9	315 (E)	311	CFOM 1: 2.0052	0.8052	343.0	34.73	89.5	-	FAILURE
	NSRTOT	7558	PHASES				CFOM 2: 1.8753	0.8295	375.7	35.37	71.8	-	
	CONVERGE			2 0 2	360,180	360	CFOM 3: 1.8422	0.7681	335.8	35.32	86.7	-	
	PROB	0.95	PERMUTED	3 15 2	45,135,225,315	135	CFOM 4: 1.8256	1.2568	787.3	35.74	88.4	-	
	NSPEC,NGEN,NANY	0,0,3	PHASES	3 1 6	45,135,225,315	111	CFOM 5: 1.8139	0.7829	309.2	35.99	86.2	-	
	FASTAN						PUBLISHED PHASES	0.9637	212.6	26.27	10.0*		
	NUMSET	32					PUB REFINED BY FASTAN	1.0116	234.4	26.80	10.9		
5	SIGMA-2		SIGMA-1	2 6 0	180, P=0.977	180	ALL 64: MAXIMUM	1.2603	787.2	41.72	90.2		
	NUMB	390	ORIGIN	3 45 0	360	360	MINIMUM	0.6650	284.2	35.20	41.7		
	NSRT	4000	DEFINING	1 3 9	315 (E)	311	CFOM 1: 2.1294	0.7970	330.6	35.20	71.0	-	
	NSRTOT	7558	PHASES				CFOM 2: 2.0256	0.7799	329.6	35.70	69.6	-	
	CONVERGE			0 12 7	360,180	180	CFOM 3: 1.9908	0.7847	346.6	35.76	68.3	-	
	PROB	0.95	PERMUTED	0 46 6	360,180	180	CFOM 4: 1.9606	0.8190	381.7	35.88	74.5	-	
	NSPEC,NGEN,NANY	2,0,2	PHASES	3 15 2	45,135,225,315	135	CFOM 5: 1.9427	1.2603	787.2	35.57	88.4	-	
	FASTAN			3 1 6	45,135,225,315	111	CFOM 6: 1.9058	0.7839	328.6	36.54	82.4	-	
	NUMSET	64					CFOM 7: 1.8966	0.7570	326.8	36.33	90.0	-	
							CFOM 8: 1.8910	0.7278	306.6	36.31	41.7	21+11 =32	1-19,21,23,26,31, 33,34,36,39-41,43, 48,73
							CFOM 9: 1.8894	0.9658	509.6	36.29	87.7	-	
							PUBLISHED PHASES	0.9637	212.6	26.27	10.0*		
							PUB REFINED BY FASTAN	1.0117	234.3	26.80	10.9		
6	SIGMA-2		SIGMA-1	2 6 0	180, P=0.977	180	ALL 64: MAXIMUM	1.1576	811.5	41.72	90.5		
	NUMB	390	ORIGIN	3 45 0	360	360	MINIMUM	0.6478	320.0	33.42	67.8		
	NSRT	7558	DEFINING	1 3 9	315 (E)	311	CFOM 1: 2.1433	0.8968	489.6	33.42	90.1	-	FAILURE
	NSRTOT	7558	PHASES				CFOM 2: 2.0463	0.8878	528.0	33.43	87.5	-	
	CONVERGE			12 26 6	45,135,225,315	222	CFOM 3: 1.8094	1.0021	556.2	36.78	89.4	-	
	PROB	0.95	PERMUTED	4 2 5	45,135,225,315	115	CFOM 4: 1.8004	0.9983	567.3	36.61	90.2	-	
	NSPEC,NGEN,NANY	0,0,3	PHASES	3 15 2	45,135,225,315	135	CFOM 5: 1.7859	1.0032	568.0	36.80	90.5	-	
	FASTAN						PUBLISHED PHASES	0.9805	214.8	17.79	7.1*		
	NUMSET	64					PUB REFINED BY FASTAN	1.0199	232.7	18.29	8.2		
7	SIGMA-2		SIGMA-1	2 6 0	180, P=0.977	180	ALL 64: MAXIMUM	1.3734	805.1	42.28	90.9		
	NUMB	390	ORIGIN	3 45 0	360	360	MINIMUM	0.7349	366.4	31.16	59.4		
	NSRT	7558	DEFINING	1 3 9	315 (E)	311	CFOM 1: 2.0145	0.7571	375.2	31.16	70.5	-	FAILURE
	NSRTOT	7558	PHASES				CFOM 2: 2.0018	0.7550	366.4	31.49	69.4	-	
	CONVERGE			12 26 6	45,135,225,315	222	CFOM 3: 1.9334	0.7530	382.3	31.82	71.9	-	
	PROB	0.95	PERMUTED	0 12 7	360,180	180	CFOM 4: 1.8508	0.8959	488.9	32.52	85.6	-	
	NSPEC,NGEN,NANY	2,0,2	PHASES	3 15 2	45,135,225,315	135	CFOM 5: 1.7937	0.9588	478.2	34.52	86.7	-	
	FASTAN			0 16 4	360,180	360	PUBLISHED PHASES	0.9805	214.8	17.79	7.1*		
	NUMSET	64					PUB REFINED BY FASTAN	1.0199	232.7	18.29	8.2		

major portion of the space required by the *MULTAN* program. Only the most reliable $\sum_2$ relationships, those with the largest κ, can be stored. The inability to store all the $\sum_2$ relationships unless their number is kept low means that not only does the user by his choice of E values define the structure to be solved but may also be forced to choose the particular set of probabilistic equations which will be used to solve it from among the complete set of existing equations. One may then ask, first, whether the partial set of equations relating the phases of the NUMB E values *has* a solution near the true solution, and second, if so, whether it is possible to arrive at it using the algorithms in *MULTAN*.

If not all the $\sum_2$ relationships can be stored, keeping the most reliable *relationships* is incompatible with being able to determine all *phases* equally well. The smaller E values will tend to have fewer of their $\sum_2$ relationships retained, and will therefore tend to be eliminated sooner during the convergence procedure and appear near the top of the convergence map. The fewer $\sum_2$ relationships retained, the worse will tend to be the determination of a given phase with the tangent formula; the tangent-formula phase shift defined above sets the limit on the accuracy possible with the terms which are included.

This is shown in Figs. 1–4 (LITH, MINA, RR, TPH respectively), where, for all the combinations of NUMB and NSRT investigated, the mean tangent formula phase shift is plotted *versus* decreasing magnitude of E, represented by code numbers denoting E magnitude ranking. Means were calculated

Table 7. AZET: $C_{21}H_{16}ClNO$, $Pca2_1$, $Z=8$, NAT=24+24=48

CASE	PARAMETERS		STARTING SET						RESULTS				MEAN	E-MAPS	
			ENANTIOMORPH FIXED BY (E)						PHASE SET	FIGURES OF MERIT			PHASE	CORRECT PEAKS	
			TYPE	h	k	l	PHI	PUB	CONSIDERED	ABS FOM	PSI ZERO	RESID	ERROR	TOTAL	PEAK NUMBERS
1	SIGMA-2		SIGMA-1	0	4	0	360, P=1.000	360	ALL 64: MAXIMUM	1.8106	477.8	33.65	72.4		
	NUMB	240		0	6	0	180, P=1.000	180	MINIMUM	1.5235	473.7	27.54	33.1		FAILURE
	NSRT	4046	ORIGIN	15	3	1	353	353	CFOM 1-57: 1.9672-	1.8064-	473.7-	27.54-	33.1-	-	
	NSRTOT	4046	DEFINING	11	3	2	225 (E)	249	2.9861	1.8106	477.3	28.46	44.0		FOR TYPICAL E-MAPS
	CONVERGE		PHASES	2	5	4	45,315	316	CFOM 58-64: 0.0000-	1.5235-	475.4-	33.00-	70.9-	-	SEE BELOW
	PROB	0.95		2	7	0	360,180	360	0.6773	1.5267	477.8	33.65	72.4		
	NSPEC,NGEN,NANY	0,0,3	PERMUTED	9	8	2	45,135,225,315	52	PUBLISHED PHASES	1.3516	255.8	20.83	13.5•	44	SEE BELOW
	FASTAN		PHASES	9	3	4	45,135,225,315	288	PUB REFINED BY FASTAN	1.8101	474.0	27.64	33.5		
	NUMSET	64													
2	SIGMA-2		SIGMA-1	0	4	0	360, P=1.000	360	ALL 32: MAXIMUM	1.8116	632.7	46.40	80.3		
	NUMB	300		0	6	0	180, P=1.000	180	MINIMUM	1.1814	490.3	35.42	39.6		FAILURE
	NSRT	4000	ORIGIN	17	3	3	23	23	CFOM 1-16: 1.9313-	1.8065-	625.2-	35.42-	39.6-	-	
	NSRTOT	7425	DEFINING	2	5	4	315 (E)	316	2.0351	1.8116	632.7	36.23	46.8		FOR TYPICAL E-MAPS
	CONVERGE		PHASES	9	3	4	45,315	288	CFOM 17-28: 1.5418-	1.5683-	614.4-	36.47-	73.0-	-	SEE BELOW
	PROB	0.95		2	7	0	360,180	360	1.6519	1.5724	618.3	37.44	76.6		
	NSPEC,NGEN,NANY	0,0,3	PERMUTED	8	2	0	360,180	180	CFOM 29-32: 0.7158-	1.1814-	490.3-	44.92-	76.9-	-	
	FASTAN		PHASES	9	8	2	45,135,225,315	52	1.0554	1.2201	532.6	46.40	80.3		
	NUMSET	32							PUBLISHED PHASES	1.3748	316.5	28.50	15.4•	46	SEE BELOW
									PUB REFINED BY FASTAN	1.8092	625.3	35.45	38.0		
3	SIGMA-2		SIGMA-1	0	4	0	360, P=1.000	360	ALL 32: MAXIMUM	1.8255	646.4	40.67	81.9		
	NUMB	300		0	6	0	180, P=1.000	180	MINIMUM	1.2828	593.0	32.01	39.3		FAILURE
	NSRT	7425	ORIGIN	17	3	3	23	23	CFOM 1-23: 1.9493-	1.8236-	642.4-	32.01-	39.3-	-	
	NSRTOT	7425	DEFINING	2	5	4	315 (E)	316	2.0715	1.8255	646.4	32.49	46.8		FOR TYPICAL E-MAPS
	CONVERGE		PHASES	7	2	0	180	180	CFOM 24-28: 1.2539-	1.5443-	631.1-	36.12-	73.7-	-	SEE BELOW
	PROB	0.95		2	7	0	360,180	360	1.2943	1.5449	632.1	36.30	74.9		
	NSPEC,NGEN,NANY	0,0,3	PERMUTED	9	8	2	45,135,225,315	52	CFOM 29-32: 0.9478-	1.2828-	593.0-	39.59-	81.5-	-	
	FASTAN		PHASES	3	1	2	45,135,225,315	67	1.1378	1.3001	598.5	40.67	81.9		
	NUMSET	32							PUBLISHED PHASES	1.3464	316.4	22.67	13.7•	46	SEE BELOW
									PUB REFINED BY FASTAN	1.8241	642.4	31.95	38.5		
4	SIGMA-2		SIGMA-1	0	4	0	360, P=1.000	360	ALL 64 NOT DEVELOPED	-	-	-	-		
	NUMB	400		0	6	0	180, P=1.000	180	STARTING SET PHASES CLOSEST TO THEIR PUBLISHED VALUES	1.8305	897.9	39.66	45.2	-	FAILURE
	NSRT	8000	ORIGIN	17	3	3	23	23							
	NSRTOT	16274	DEFINING	9	3	4	315 (E)	288							
	CONVERGE		PHASES	7	2	0	180	180	PUBLISHED PHASES	1.3937	445.7	31.16	14.0•	24+23 =47	1-19,21-32,34-37, 39,40,42,43,46,48, 50,56,58,60,61,67 (1,2=Cl)
	PROB	0.95		2	7	0	360,180	360							
	NSPEC,NGEN,NANY	2,0,2	PERMUTED	8	2	0	360,180	180							
	FASTAN		PHASES	2	5	4	45,135,225,315	316							
	NUMSET	64		9	8	2	45,135,225,315	52	PUB REFINED BY FASTAN	1.8295	895.6	39.43	39.1		

CASE	RESULTS				MEAN	TYPICAL E-MAPS	
						CORRECT PEAKS	
	PHASE SET CONSIDERED	ABS FOM	PSI ZERO	RESID	PHASE ERROR	TOTAL	PEAK NUMBERS
1	CFOM 1: 2.9861	1.8098	473.7	27.61	37.0	14+13=27	1-5,7,9,19,21,24-27,29,37,43,44,46,52-54,60,64,65,70,84,90 (1,2=Cl)
	CFOM 2: 2.9814	1.8095	473.7	27.54	33.7	16+14=30	1-6,8,11,15,20,21,25,30,33,35,36,38,46-49,53,58,72,73,75,77,80,90,96 (1,4=Cl)
	CFOM 3: 2.9127	1.8100	473.9	27.70	38.5	14+12=26	1-4,6,7,11,14,20,24,25,27,31,32,35,37,39,50,52,53,60,62,70,76,77,95 (1,2=Cl)
	PUBLISHED PHASES	1.3516	255.8	20.83	13.5•	24+20=44	1-15,17-19,22-24,26-29,31-33,37,39,41,44,46-48,50,51,54-56,58,67,69,93 (1,2=Cl)
2	CFOM 1: 2.0351	1.8105	627.4	35.42	42.7	17+10=27	1-4,7,8,10,11,19,25,26,28,31-33,39,46,52,59,60,62,65,66,72,73,94,95 (2,3=Cl)
	CFOM 2: 2.0336	1.8111	627.5	35.43	43.5	14+ 9=23	1,2,4,7,8,11,17,18,26,30,33.37,40,42,45,49,52,55,56,63,68,93,95 (1,8=Cl)
	CFOM 3: 2.0318	1.8109	627.7	35.44	43.5	11+ 9=20	1,2,4,6,8,9,13,16,18,23,32,35,36,52,70,75,84,85,87,93 (2,18=Cl)
	PUBLISHED PHASES	1.3748	316.5	28.50	15.4•	23+23=46	1-12,14,16-18,20,22,30,33,35,37,40-45,47,49,50,53,56-59,82,86,96 (1,2=Cl)
3	CFOM 1: 2.0715	1.8242	642.4	32.01	41.8	14+13=27	1-3,5,7,11,14,15,19,23,25,28,31,32,36,38,43,44,49,54,57,67,69,71,72,84,85 (2,3=Cl)
	CFOM 2: 2.0633	1.8246	642.8	32.03	46.2	14+10=24	1-6,13,17,21,33-35,39,40,47,48,53,58,59,70-72,76,93 (2,4=Cl)
	CFOM 3: 2.0607	1.8249	642.9	32.04	46.8	14+12=26	1,2,4,5,8,9,12,14,15,17,30,33,38,45,51-53,56,61,62,78,82,92-95 (2,5=Cl)
	PUBLISHED PHASES	1.3464	316.4	22.67	13.7•	23+23=46	1-12,14,16-18,20,22-30,33,35,37.40-45,47,49,50,53,56-59,82,86,96 (1,2=Cl)
4	PUBLISHED PHASES	1.3937	445.7	31.16	14.0•	24+23=47	1-19,21-32,34-37,39,40,42,43,46,48,50,56,58,60,61,67 (1,2=Cl)

for overlapping batches of 20 reflexions. For sets of complete $\sum_2$ relationships, the phase shift is small and relatively uniform with decreasing $|E|$ and up the convergence map. As the number of $\sum_2$ relationships is restricted, the phase shift tends to increase with decreasing $|E|$ and towards the top of the convergence map. The phases at the top of the convergence map, although they may be far from correct, are recycled and used in refining already determined phases at the bottom of the map. It is not hard to see how this process could lead to divergence of the entire phase set from approximately correct values, even though the initial phase development may have been quite accurate. On the other hand, if the initial phase development is only moderately good, having sufficient $\sum_2$ relationships may allow a better chance of refinement to quite accurate phases, with the phase errors distributed roughly evenly among all the reflexions.

The user's aim should therefore be, according to the size of the structure to be solved and the computer storage available, to *use only as many E values as necessary but as many* $\sum_2$ *relationships as possible*, preferably all. This is the only way with the algorithms presently in the *MULTAN* program to ensure that the ratio of $\sum_2$ relationships to E values is as uniform as possible and thus that the tangent formula phase shift is as low and as uniform as possible. One should avoid having convergence maps with long narrow tails at the top, since adding to the number of E values without being able to use the equations necessary to determine their phases at all well may not improve the definition of the structure in the Fourier map but only increase the noise, and actually yield diminishing returns. The various cases in Tables 3–6 (LITH, MINA, RR, TPH respectively) show the increased *likelihood* of success when the set of $\sum_2$ relationships used is complete or at least reasonably large, which can be achieved either by increasing NSRT or by decreasing NUMB, when possible.

This approach is fundamental to the intelligent use of *MULTAN*. Structures may be and quite often are solved when the parameters NUMB and NSRT are not ideal, but in case of failure, improving the choice of these two parameters may be the best first change to make, and they should be kept at sensible values to form a reliable basis for any other attempts at changed or improved tactics if these are required.

Because the convergence map differs, in general, for each different choice of parameters, the path of phase determination may in one case encounter a favourable and in another an unfavourable sequence of invariants. Thus a number of attempts at solution may often be necessary, and the user should not be discouraged too easily. Even with a fairly easily solvable structure like MINA, an 'unlucky' convergence map led to failure in the initial attempt (Table 4, case 2) while all other choices of parameters yielded the solution. More difficult structures, like RR (Table 5) and TPH (Table 6), clearly present more formidable obstacles to be got round, and most attempts at solving them failed, but even these two structures eventually yielded to persistence.

Other tactics: acceptance of $\sum_1$ indications

Occasionally a $\sum_1$ phase indication of high probability is incorrect. For E-set 1 of INOS, the $\sum_1$ formula, with two contributors among the largest 260 E values, indicates that the phase of 008 is 360° with probability 0·999, but this is incorrect, and is an obstacle to correct phase determination (Table 2, case 1). If PROB is raised to 1·0 (Table 2, case 2), this false $\sum_1$ indication is not accepted, and 008 no longer appears in the starting set. $\sum_2$ invariants of 180° still prevent a completely successful phase determination, but now the set

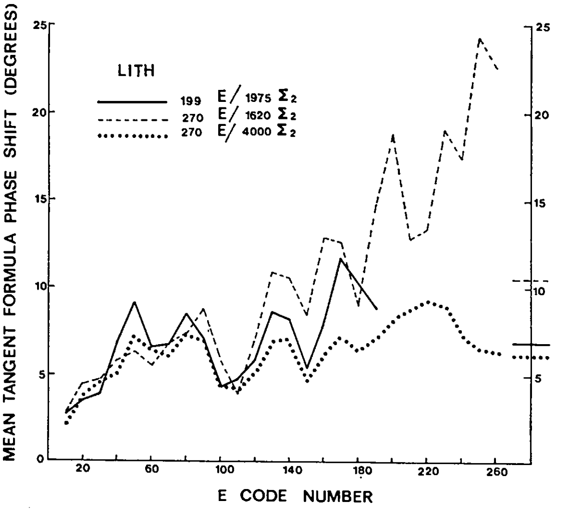

Fig. 1. LITH: mean tangent formula phase shift *vs.* decreasing $|E|$.

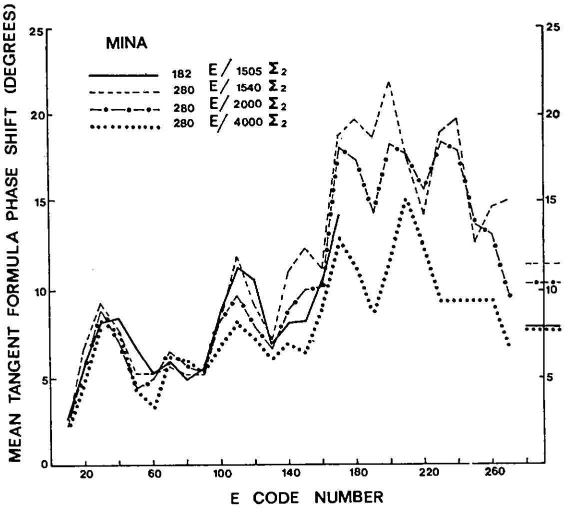

Fig. 2. MINA: mean tangent formula phase shift *vs.* decreasing $|E|$.

with the highest CFOM gives an image of the structure in the E map, displaced from its correct position by an average $\Delta y = -0{\cdot}33$. This must be counted a partial success, since there are several procedures for finding the correct positions of molecules once their orientations are known. Two similar recent instances which may be cited are those of avicennin ($C_{20}H_{20}O_4$, $P\bar{1}$, $Z=2$) (Ting & Marsh, 1974), with an incorrect $\sum_1$ indication of probability 0·98, and of diphenyl-Δ^2-pyrazoline ($C_{15}H_{14}N_2$, $P2_1/c$, $Z=4$) (Duffin, 1968; Gassman & Zechmeister, 1972), with an incorrect $\sum_1$ indication of probability 0·97. In both these cases the E maps showed multiple shifted images of the structure.

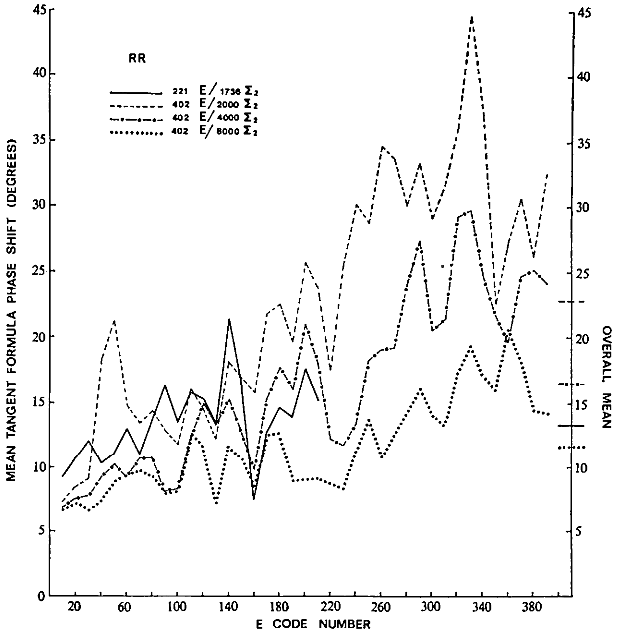

Fig. 3. RR: mean tangent formula phase shift *vs.* decreasing $|E|$.

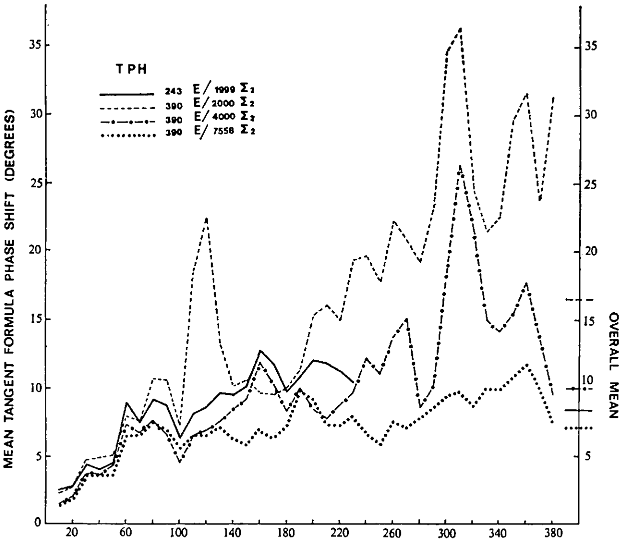

Fig. 4. TPH: mean tangent formula phase shift *vs.* decreasing $|E|$.

There is of course no simple way of knowing whether a $\sum_1$ indication is correct or not. If examination of the convergence map shows that a phase accepted from a $\sum_1$ indication enters in an important way into the early stages of phase determination, then the tactical choice of PROB can be a significant one. Increasing PROB and not accepting some $\sum_1$ indication can help if the indication is wrong, but may hurt if it is right; one can only attempt the phase determination both ways and see.

Other tactics: increasing the starting set

An evident tactical change in case of failure is to take more unknown phases, to be permuted, into the starting set. This was successful with INOS (Table 2, cases 5 and 6 *vs.* cases 3 and 4), LITH (Table 3, case 3 *vs.* case 2, case 5 *vs.* case 4), MINA (Table 4, case 3 *vs.* case 2), and, to a lesser extent, TPH (Table 6, case 5 *vs.* case 4). At first sight, the increased chance of success might seem to be only a matter of increasing the amount of phase information at least approximately known at the outset, and increasing the constraints on phase development, since the starting-set phases are kept fixed until the final cycles of phase determination and refinement. Very often this is indeed what happens, but in general the phase determining process with *MULTAN* can be rather more complex than that.

It is *not* true that the starting set with phases closest to their true values is always the one which develops into the correct solution. This is because the invariants encountered in the convergence map, even toward the beginning, may differ substantially from zero. Multisolution numerical phase development makes a certain allowance for this (albeit in a somewhat haphazard way), since permutation of the values assigned to unknown *phases* in the starting set amounts to the exploration of a range of possibilities for the structure *invariants*. An accurate starting set may diverge from the correct solution, while a fairly poor starting set, because of the distribution of invariants encountered in the course of phase development, may finally converge to the correct set of phases. This makes the permutation strategy more powerful than is at first evident, and happens often enough to make it the strongest single argument in favour of the numerical multisolution approach as opposed to symbolic addition (Karle & Karle, 1966).

The reasons why adding more unknown phases to the starting set increases the chances of success should now be clear. However, any change in NSPEC, NGEN, NANY will in general alter the convergence map and thus change the path of phase development which it determines. This means that occasionally a phase determination with *fewer* reflexions in the starting set will succeed while one with *more* reflexions will fail, as happened with RR (Table 5, case 5 *vs.* case 6).

The previous discussion of the choice of NUMB and NSRT is relevant as well to a consideration of practical limitations on increasing the number of unknown phases in the starting set. The time necessary to run the *MULTAN* program consists of a relatively short block used in finding the $\sum_2$ relationships and performing the convergence mapping plus a generally much longer block used in developing all phase sets with *FASTAN*. The length of this second block is proportional to NUMSET, and the time needed for each phase set is proportional to NSRT. Therefore, if NSRT is to be a large fraction of NSRTOT, as it should, the choice of a smaller NUMB allows a greater number of unknown reflexions to be added to the starting set before practical limitations on computing time are reached. Choice of a *smaller* NUMB thus makes *MULTAN* a *more* powerful program in practical terms.

NUMSET increases by a factor of four for each general reflexion and two for each special reflexion added to the starting set, so the user can also take advantage of special circumstances to increase the size of the starting set while keeping NUMSET within reasonable bounds. When a phase determination fails for a structure like TPH, space group $C222_1$, having large numbers of special reflexions with high E values, the user can specify that reflexions added to the starting set be special. This is shown in Table 6, where case 4, NANY = 3, NUMSET = 32, is a failure, whereas case 5, NSPEC = 2, NANY = 2, NUMSET = 64, gives one phase set for which most of the asymmetric unit is easily recognizable in the E map, and from which the complete structure is readily found. This case incidentally illustrates the practical importance of being able rapidly to compute, examine, and interpret large numbers of E maps (Declercq, Germain, Main & Woolfson, 1973; Koch, 1974). The best set in case 5 still has significantly large phase errors in it (mean phase error 41·7°), and its astonishingly low ABS FOM places it eight in CFOM ranking, for which reasons it could easily have been overlooked if computing and examining E maps were a lengthy and tedious operation.

The difficulty in solving RR clearly illustrates the practical limits on the tactical choice of increasing the starting set which arise from the strategy of independently permuting the phases of the unknown starting reflexions. Because there are very few special reflexions with large E values for RR, for sensible phase development not only must all the unknown phases in the starting set be general (each new one taken multiplying NUMSET by a factor of four), but the origin must be defined using three general reflexions as well. One of these can be assigned an arbitrary phase, the second can have its phase restricted to define the origin and enantiomorph, but the third must be given *two* values to ensure that one of the values corresponds to the enantiomorph defined by the second phase. Thus, with the origin defined in this manner, to have just three unknown phases in the starting set for RR requires the development of 128 phase sets, which is a large computing job. If such an attempt fails, as in fact it did (Table 5, cases 2, 4, 6, 8), a decision to add one more unknown phase would mean having to compute 512 sets of phases, a number impracticable for most users.

An improved strategy which has been developed to overcome the limitations of independent permutation of unknown phases is the 'magic integer' technique (White & Woolfson, 1975; Declercq, Germain & Woolfson, 1975). This is a method of correlating the values of the unknown phases with $\sum_2$ relationships and then choosing only those sets of values which best satisfy the requirement that these $\sum_2$ invariants be near zero. When only the most likely combinations of the unknown phases are selected in this way, a much larger number of phases, say 10–30 or more, can be used while keeping NUMSET about the same as it would be using only three or four independently permuted phases in the present version of *MULTAN*. The 'magic integers' strategy is therefore in general more powerful than the independent permutation strategy.

The inadequacy of the tangent formula: an ultimate limitation on the *MULTAN* strategy

Examination of the results of attempts to solve the structure of AZET with *MULTAN* (Table 7) shows that to have any hope of a successful phase determination we require a third property of the defined struc-

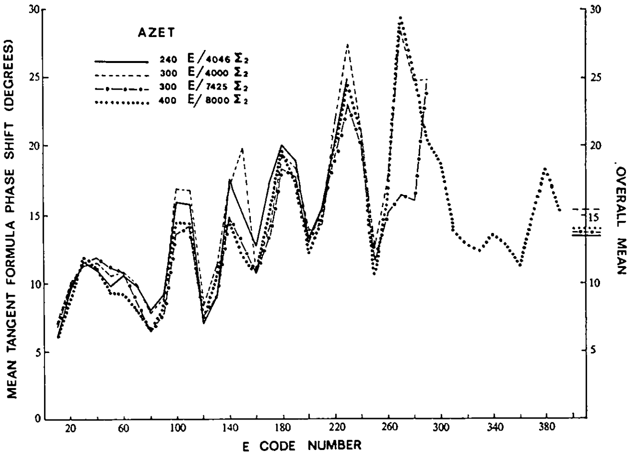

Fig. 5. AZET: mean tangent formula phase shift *vs.* decreasing $|E|$.

ture, namely, stability of the true phase set under tangent formula refinement. For AZET, the first two requirements are met. The structures defined with the 240, 300, and 400 largest E values (Table 7, cases 1, 2, and 4 respectively) are all readily interpretable in terms of the actual structure. Likewise, the mean tangent-formula phase shift for the published phases (Table 7, cases 1–4, and Fig. 5), which sets an absolute limit on the possible accuracy of phase determination, is quite low.

In practice, however, the actual limit of accuracy to be expected when phases are determined by *MULTAN* is given by a quantity listed for each case in the tables, the mean phase error resulting when the published phases are refined by *FASTAN*. This is borne out by examining all the successful phase determinations for the five structures previously discussed. For these structures, whether the phase determination was successful or not, in every case the mean phase error for PUB refined by *FASTAN* is only slightly higher than the mean tangent formula phase shift for PUB, and the figures of merit for the former remain close to those for the latter.

For AZET, however, this is not so. The published phases for AZET are simply unstable under the operations of *FASTAN*, and diverge from their true values to a mean phase error much higher than the mean tangent formula phase shift. The results from *MULTAN* are in fact as good as could be expected, but the best phase sets give very poor, fragmentary, and unrecognizable representations of the structure (Table 7, cases 1, 2, 3, typical E maps). Similar divergence of published phases refined by *FASTAN* is reported for the potassium salt of alborixin ($C_{48}H_{83}O_{14}K$, $P2_1$, $Z=2$) (Alléaume, Busetta, Farges, Gachon, Kergomard & Staron, 1975; Busetta, 1976). (Note that this behaviour cannot be ascribed merely to the presence of heavy atoms, since *MULTAN* routinely solves similar structures, sometimes with the heavy atom a larger fraction of the total scattering matter, as, for example, 2,4,6-trimethyldiphenyl sulphone ($C_{15}H_{16}O_2S$, $Pn2_1a$, $Z=8$) (Chawdhury, 1976). Note also that the two chlorine atoms in AZET do *not* in general show up as the highest two peaks in the E maps.)

There thus exist crystal structures which it is *impossible* to solve with *MULTAN*. Conceivably this is due to poor or not universally applicable tactics in *FASTAN*, but more likely it is because the fundamental *MULTAN* strategy, use of the tangent formula for phase determination and refinement, is inadequate (*cf.* Gassman & Zechmeister, 1972). To revise *MULTAN* we must seek other phase-determining strategies which contain constraints to prevent instability of the sort seen with AZET, which can be programmed and made automatic, and which we would like to be not excessively costly in computer time compared with the rapid and convenient tangent formula.

Future strategies

Several more powerful general phase-determining formulae have been proposed. Extended tangent formulae, containing quartet structure invariants, of the form

$$\tan \varphi_{\mathbf{h}} = \frac{\sum_{\mathbf{k}} w_{\mathbf{hk}} \sin(\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}}) + \sum_{\mathbf{k}} \sum_{\mathbf{l}} w_{\mathbf{hkl}} \sin(\varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} + \varphi_{\mathbf{h}-\mathbf{k}-\mathbf{l}})}{\sum_{\mathbf{k}} w_{\mathbf{hk}} \cos(\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}}) + \sum_{\mathbf{k}} \sum_{\mathbf{l}} w_{\mathbf{hkl}} \cos(\varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} + \varphi_{\mathbf{h}-\mathbf{k}-\mathbf{l}})}$$

can be derived from (*a*) requiring that the integral of the square of the difference between the squared defined structure and the defined structure, properly scaled, be a minimum (Allegra & Colombo, 1974); (*b*) considering the conditional probability distribution of the quartet structure invariant given the magnitudes of the four E values (Hauptman, 1975*a*,*b*); or (*c*) applying the maximum determinant rule (Tsoucaris, 1970) to an order-4 Karle–Hauptman matrix (Main, 1975).

The maximum determinant rule can of course be applied to matrices of increasingly high order, and expressed as a general *n*th-order tangent formula (de Rango, Tsoucaris & Zelwer, 1974; Mauguen, de Rango & Tsoucaris, 1973). Higher-order phase-determining formulae of quite general types can also be formulated conveniently by considering the process of phase correction in direct space (Gassman & Zechmeister, 1972). The approach in all these cases is to develop more powerful phase relationships among the large E values defining the structure.

A different tack, however, is strongly suggested by the behaviour of the figures of merit when the published phases are refined by *FASTAN*, compared with the values for the published phases themselves. For both AZET and the potassium salt of alborixin (Busetta, 1976) this leads to a most unusually large absolute figure of merit, coupled with a greatly increased ψ_0 figure of merit. In other words, application of the tangent formula drives the true phases to values which make the $\sum_2$ relationships among the largest E values very consistent, the invariants much more narrowly distributed about zero than they in fact are, at the expense of failing to satisfy nearly as well the Sayre's equations for the smallest E values, which are used only to calculate ψ_0 but not for phase development itself. The most promising future strategies for *MULTAN* might thus be methods which could incorporate the smallest E values into the phase development procedure in a way which would constrain the phases of the largest E values. This would use more of the complete data set, in such a way as to impose on the defined structure the additional requirement of a correspondence to properties of the actual structure more strict and of a different kind than simply a reasonably good 1:1 match of the largest peaks.

The basis of one such approach, the use of quartets with probably negative cosines, has recently appeared (Hauptman, 1975*a*,*b*). From a consideration of conditional probability distributions involving seven structure factors, Hauptman has shown that a quartet structure invariant may be distributed about any value between 0° and 180°, depending on the magnitudes of both the four E values of the quartet and the three E values of the cross terms. In particular, the larger the former and the smaller the latter, the more narrowly the quartet is distributed about 180°, in sharp contrast to the prediction of a distribution with a maximum always at 0° when only the four magnitudes of the quartet are considered. Taking into account the smallest E values thus clearly does change the relations which the phases of the largest E values must satisfy, and would constrain the phase development if these 'negative quartets' were incorporated into the procedure. It seems likely that in the immediate future at least some improvements in phase development procedures will occur along these lines.

I would like to thank Professor Michael Woolfson, Dr Peter Main, and Dr Gabriel Germain for their interest and encouragement in this work. Data for analysis were kindly provided for INOS by Dr D. A. Langs, Medical Foundation of Buffalo, and Professor H. C. Freeman, University of Sydney; for LITH by Dr S. K. Arora and Professor R. B. Bates, University of Arizona; for MINA by Dr F. Plana, University of Barcelona; for RR by Dr H. Henke, University of Karlsruhe; for TPH by Dr A. Hoekstra, University of Groningen; and for AZET by Mr A. Colens, University of Louvain. The financial support of the Science Research Council is gratefully acknowledged.

References

Alléaume, M., Busetta, B., Farges, C., Gachon, P., Kergomard, A. & Staron, T. (1975). *Chem. Commun.* pp. 411–412.

Allegra, G. & Colombo, A. (1974). *Acta Cryst.* A**30**, 727–729.

Arora, S. K., Germain, G. & Declerq, J. P. (1976). *Acta Cryst.* B**32**, 415–419.

Busetta, B. (1976). *Acta Cryst.* A**32**, 139–143.

Chawdhury. S. A. (1976). *Acta Cryst.* B**32**, 1069–1073.

Colens, A., Declercq, J. P., Germain, G., Putzeys, J. P. & Van Meerssche, M. (1974). *Cryst. Struct. Commun.* **3**, 119–122.

Debye, P. (1915). *Ann. Phys.* (*Leipzig*), **46**, 809–823.

Declercq, J. P., Germain, G. & Henke, H. (1973). *Cryst. Struct. Commun.* **2**, 405–409.

Declercq, J. P., Germain, G., Main, P. & Woolfson, M. M. (1973). *Acta Cryst.* A**29**, 231–234.

Declercq, J. P., Germain, G. & Woolfson, M. M. (1975). *Acta Cryst.* A**31**, 367–372.

Duffin, B. (1968). *Acta Cryst.* B**24**, 1256–1261.

Font-Altaba, M., Miravitlles, C., Brianso, J. L., Plana, F. & Solans, X. (1976). *Acta Cryst.* To be published.

Freeman, H. C., Langs, D. A., Nockolds, C. E. & Oh, Y. L. (1976). In preparation.

Gassman, J. & Zechmeister, K. (1972). *Acta Cryst.* A**28**, 270–280.

Germain, G., Main, P. & Woolfson, M. M. (1970). *Acta Cryst.* B**26**, 274–285.

Germain, G., Main, P. & Woolfson, M. M. (1971). *Acta Cryst.* A**27**, 368–376.

Germain, G. & Woolfson, M. M. (1968). *Acta Cryst.* B**24**, 91–96.

Hauptman, H. (1975*a*). *Acta Cryst.* A**31**, 671–679.

Hauptman, H. (1975*b*). *Acta Cryst.* A**31**, 680–687.

Hoekstra, A., Vos, A., Braun, P. B. & Hornstra, J. (1975). *Acta Cryst.* B**31**, 1708–1715.

Karle, J., Hauptman, H. & Christ, C. L. (1958). *Acta Cryst.* **11**, 757–761.

Karle, J. & Karle, I. L. (1966). *Acta Cryst.* **21**, 849–859.

Koch, M. H. J. (1974). *Acta Cryst.* B**30**, 67–70.

Main, P. (1975). Private communication.

Main, P., Woolfson, M. M., Lessinger, L., Germain, G. & Declercq, J. P. (1974). *MULTAN* 74, *A System of Computer Programmes for the Automatic Solution of Crystal Structures from X-Ray Diffraction Data.*

Mauguen, Y., de Rango, C. & Tsoucaris, G. (1973). *Acta Cryst.* A**29**, 574–575.

Mo, F., Hjortås, J. & Svinning, T. (1973). *Acta Cryst.* A**29**, 358–362.

Rango, C. de, Tsoucaris, G. & Zelwer, C. (1974). *Acta Cryst.* A**30**, 342–353.

Roberts, P. J., Brauer, D. J., Tsay, Y. -H. & Krüger, C. (1974). *Acta Cryst.* B**30**, 2673–2678.

Ting, F. L. & Marsh, R. E. (1974). *Acta Cryst.* B**30**, 1570–1575.

Tsoucaris, G. (1970). *Acta Cryst.* A**26**, 492–499.

White, P. S. & Woolfson, M. M. (1975). *Acta Cryst.* A**31**, 53–56.