CROATICA CHEMICA ACTA CCACAA **72** (2–3) 519–529 (1999)

ISSN-0011-1643
CCA-2594 *Original Scientific Paper*

Ab initio Direct Phasing in Macromolecular Crystallography: an Application of the *Z*-test*

Maria Cristina Burla,[a] *Carmelo Giacovazzo,*[b,**] *Doriano Lamba,*[c,d] *Giampiero Polidori,*[a] *and Giovanni Ughetto*[d]

[a] *Dipartimento di Scienze della Terra, Universita' di Perugia, I-06100 Perugia, Italy*

[b] *Dipartimento Geomineralogico, Universita' di Bari, Via Orabona 4, I-70125 Bari, Italy*

[c] *International Centre for Genetic Engineering and Biotechnology, Area Science Park, Padriciano 99, I-34012 Trieste, Italy*

[d] *Ist. di Strutturistica Chimica »G. Giacomello«, Area della Ricerca di Roma – CNR, C.P.10, I-00016 Monterotondo Stazione (Roma), Italy*

Received November 3, 1998; accepted March 17, 1999

The *z*-criterion has been recently formulated as a tool for judging about the *ab initio* solvability of a crystal structure *via* direct methods. The criterion is reconsidered to take into account the recent powerful techniques of phase refinement in direct and reciprocal space. A report is made on a medium size crystal structure, recently solved by SIR97 by the pure application of the tangent formula.

Key words: macromolecular crystallography, phase determination.

INTRODUCTION

It is well known that in practice direct methods have solved the phase problem for small molecules.[1] The tangent formula was the main tool for

---

* Dedicated to Professor Boris Kamenar on the occasion of his 70th birthday.

** Author to whom correspondence should be addressed. (E-mail: c.giacovazzo@area.ba.cnr.it)

this success: crystal structures with up to 100 atoms in the asymmetric unit could be routinely solved by the application of the so called multisolution phasing approach,[2,3] but very few structures with more than 200 atoms have been solved by a pure tangent approach.

The tremendous increase in the computer power has recently made small protein molecules solvable *ab initio*: the goal was first attained by Shake and Bake,[4] later on by Half-bake and more recently by SIR99.[5,6] The three programs have different phasing strategies: for the first two, each trial solution is repeatedly cycled both in real and reciprocal space, and structure (or phase) refinement is performed in each space. The third program adopted a sophisticated real space refinement just after the application of the tangent formula. The mathematical techniques used by the three programs could make feasible the *ab initio* solution of crystal structures up to 1000 atoms in the asymmetric unit.

The question is now: which is the limiting size of the crystal structures solvable only by the tangent formula? An answer to this question may be provided by making suitable use of the z-criterion.[7] We recall here its main features.

In accordance with the tangent formula, the distribution of ϕ_h when r pairs (ϕ_k, $\phi_{h\text{-}k}$) are known is given by

$$P(\phi_h|...) = [2\pi I_0(\alpha_h)]^{-1} \exp[\alpha_h \cos(\phi_h - \vartheta_h)], \tag{1}$$

where ϑ_h, the most probable value of ϕ_h, is defined by

$$\tan \vartheta_h = \frac{\sum_{j=1}^{r} G_j \sin(\phi_{k_j} + \phi_{h-k_j})}{\sum_{j=1}^{r} G_j \cos(\phi_{k_j} + \phi_{h-k_j})} = \frac{T_h}{B_h}, \tag{2}$$

and

$$\alpha_h = [T_h^2 + B_h^2]^{1/2} \tag{3}$$

is the reliability parameter. Small values of α suggest that the relation $\phi_h \approx \vartheta_h$ is unreliable. Hence, α_h is distributed according to[8]

$$P(\alpha_h) \approx N(\alpha_h; \langle \alpha_h \rangle, \sigma^2_{\alpha_h}), \tag{4}$$

where N denotes the normal distribution, and

$$\langle \alpha_h \rangle = \sum_{j=1}^{r} G_j D_1(G_j), \tag{5}$$

$$\sigma^2_{\alpha_h} \approx \frac{1}{2}\sum_{j=1}^{r} G_j^2[1 + D_2(G_j) - 2D_1^2(G_j)] \ . \quad (6)$$

$D_i(x) = I_i(x) / I_0(x)$ is the ratio of two modified Bessel distributions. It was proposed that,[7] in the absence of phase information, the ratio

$$z_h = \langle \alpha_h \rangle / \sigma_{\alpha_h}$$

could be considered as a »signal-to-noise ratio«. Accordingly, diffraction data for which sufficiently high values of z may be calculated for most of the strong reflections constitute a good premise for the successful application of Direct Methods. For proteins of an average size, the following relationships are typical:

$$\langle \alpha_h \rangle \approx \sum_{j=1}^{r} G_j^2/2, \quad (7)$$

$$\sigma^2_{\alpha_h} \approx \sum_{j=1}^{r} G_j^2/2 \quad (8)$$

and consequently

$$\langle \alpha_h \rangle \approx \sigma^2_{\alpha_h}$$

and

$$z_h \approx \langle \alpha_h \rangle^{1/2}.$$

It may be concluded that if $\langle \alpha_h \rangle < 1$, then the signal is smaller than the noise and the tangent formula should work unreliably. The z-criterion suggests that a crystal structure is solvable *via* the tangent formula if, for a sufficiently large set of normalized structure factors,

$$z > T$$

where T is the threshold that was suggested to be close to three (2,3).

The z-criterion is an ultimate tool when only direct methods are used. However, the modern direct methods program associate the phase refinement in reciprocal space (the tangent formula, or similar methods) with the refinement in real space, so creating a more effective phasing procedure, potentially able to solve crystal structures not solvable by the pure tangent formula.

To check the limits of the pure tangent procedures, we calculated the $P(z)$ distribution for most of the structures solved by Shake and Bake and by Half-bake.

In Table I, we have selected seven of them.

In Figure 1 the corresponding $P(z)$ distributions are shown. Not all the structures satisfy the z-test for $T > 3$: the most difficult structures seem to

TABLE I

Crystal data of the test structures

Structure code	Space group	Cell parameters	Chemical content	Non-H atoms (asym. unit)
APP[9]	$C2$	a = 34.18, b = 32.92, c = 28.44; α = 90.00, β = 105.3, γ = 90.00	C, 760; N, 212; O, 232; Zn, 4	302
BPTI[10]	$P2_12_12_1$	a = 74.1, b = 23.4, c = 28.9; α = 90.00, β = 93.03, γ = 90.00	C, 1156; N, 336; O, 596; S, 36; H, 2296	531
CRAMB[11]	$P2_1$	a = 40.763, b = 18.492, c = 22.333; α = 90.00, β = 90.61, γ = 90.00	C, 406; N, 110; O, 131; S, 12; H, 776	329
GRAM086[12]	$P2_12_12_1$	a = 31.595, b = 32.369, c = 24.219; α = 90.00, β = 90.00, γ = 90.00	C, 912; N, 160; O, 196; H, 1480	317
RUBR_DV[13]	$P2_1$	a = 19.97, b = 41.45, c = 24.41; α = 90.00, β = 108.30, γ = 90.00	C, 486; N, 114; F, 2; O, 374; S, 12; H, 1168	494
TOXII[14]	$P2_12_12_1$	a = 45.94, b = 40.68, c = 29.93; α = 90.00, β = 90.00, γ = 90.00	C, 1240; N, 340; O, 764; S, 32; H, 2812	594
VMN2[15]	$P4_32_12$	a = 28.45, b = 28.45, c = 65.84; α = 90.00, β = 90.00, γ = 90.00	C, 1072; N, 144; O, 763; Cl, 64; H, 1264	255

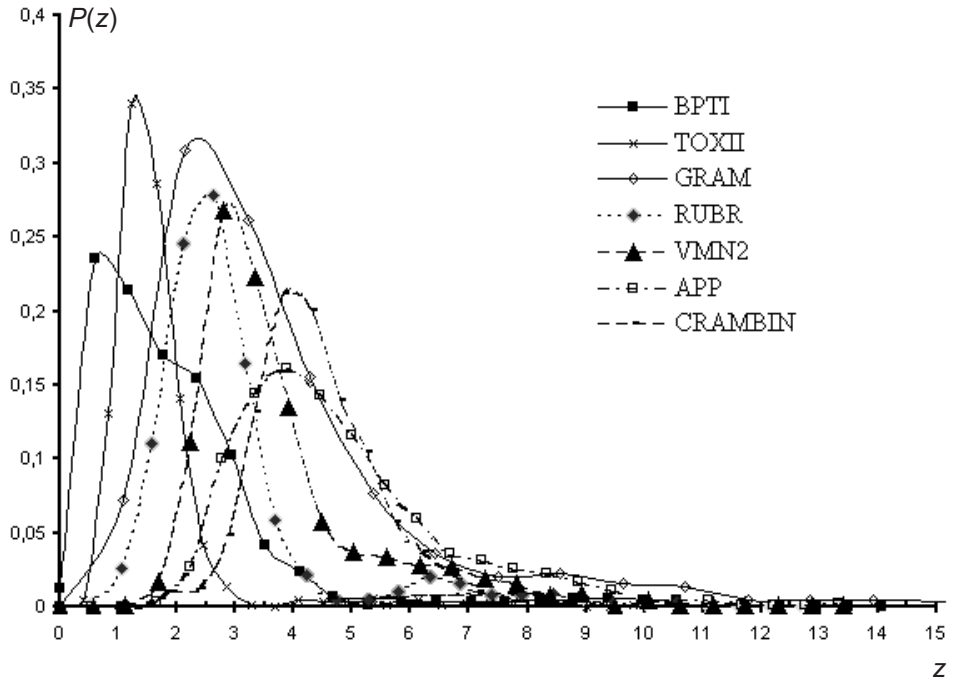

Figure 1. The $P(z)$ distributions for the seven test structures.

be BPTI and TOXII, favourable *z*-distributions are obtained for Crambin and APP.

We have introduced the *z*-test as a routine step of SIR97.[16] Experimental data for which the *z*-test is not satisfied are considered not suitable for the tangent formula application. When the experimental data of PROFL, an unknown crystal structure with 220 non-hydrogen atoms in the asymmetric unit, were tested, the $P(z)$ distribution in Figure 2 was obtained, which suggested that PROFL could be solved by applying the pure tangent formula. The approach followed for structure solution and refinement and the interesting packing properties of PROFL will be described and discussed.

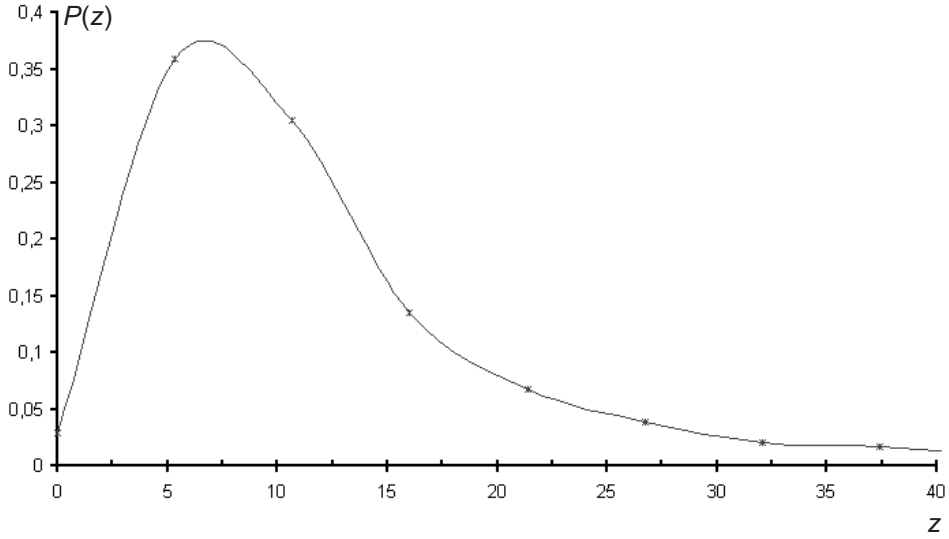

Figure 2. The $P(z)$ distributions for PROFL.

RESULTS AND DISCUSSION

The acridines derive much of their biological interest from the fact that they often show mutagenic and anti-tumour properties.[17] These compounds interact with nucleic acids; it is believed that these interactions are intimately related to the biological properties.

The study of acridines-nucleic-acid interactions has been concerned mostly with the intercalative mode of binding of planar chromophores between base pairs in double-helical DNA and RNA.[18,19] However, these mole-

cules are also known to bind to single-stranded nucleic acids such as synthethic polyribonucleotides and naturally occurring tRNA.[20,21] The exact nature of that binding is not well understood at present because of the conformational flexibility of RNA.

Proflavine, which is perhaps the most extensively studied of all acridine–nucleic acid systems has been shown to bind to both single and double stranded DNA and RNA by intercalation and external ionic binding between aminoacridine cations and charged phosphates.[22]

In view of the large number of important biological phenomena associated with intercalation, several X-ray crystal structures of self-complementary dinucleoside phosphates in complexes with acridines and other intercalating agents have been carried out.[23] In these studies, dinucleoside phosphates often crystallise as miniature double helices, however no intercalation complexes of ApU and GpC have been obtained. The pseudo-intercalation observed in the crystal structure of the complex of 9-aminoacridine with ApU is in agreement with solution studies which indicate that intercalation preferentially occurs at pyrimidine-(3',5')-purine sequences.[24]

Here, we report an attempt to crystallise and to determine the crystal and molecular structure of the aminoacridine proflavine (PROFL) in complex with the self-complementary dinucleoside phosphate adenylyl-3',5'-uridine (ApU).

A solution of 5 mg of ApU in 0.25 ml of a 100 mM phosphate buffer at pH = 6.6 was prepared. A 4.16 mg sample of proflavine hemi sulphate was dissolved in this solution, followed by addition of 0.1 ml of isopropanol. The vial was sealed and the crystals suitable for diffraction analysis appeared within a week. These crystals were unstable in the absence of the mother liquor and had to be mounted in a sealed capillary.

Lattice constants were determined and intensity data were collected on a Syntex $P2_1$ diffractometer (graphite-monochromated Cu-Kα radiation). The crystal data are a = 13.118(2) Å, b = 21.674(5) Å, c = 30.799(9) Å, α = 75.55(2)°, β = 85.19(2)°, γ = 86.62(1)°, space group $P\bar{1}$, d_m = 1.398 g cm^{-3}.

Lorentz, polarisation, but no absorption nor extinction corrections were applied. Of the 22852 unique reflections $F_o \geq 0.0$, 13487 were considered observed $F_o \geq 4.0\ \sigma(F_o)$.

The standard method for solving crystal structures of dinucleotide phosphates has been to locate the phosphorus atom by means of resolution difference techniques.[25] Many attempts to apply this technique to this structure were unsuccessful. The structure was eventually solved by direct methods using the SIR97 suite and refined,[16] based on F, using the SHELX–97 package.[26] To our surprise, determination of the crystal and molecular structure revealed only the presence of the proflavine, phosphate

and water moieties. Thus far, 11 proflavine molecules, 7 phosphate anions and 21 water molecules have been located in the asymmetric unit, with no indication of the presence of the self-complementary dinucleoside phosphate adenylyl-3',5'-uridine moiety.

This is a highly solvated structure. Lack of strong interactions among some of the solvent molecules makes them highly disordered. Therefore, the locations of these solvent molecules were difficult to determine and were sometimes ambiguous. A combination of least-squares refinement and difference Fourier syntheses was carried out step by step to determine unequivocally 21 positions occupied by ordered water molecules and so to complete the structure.

No attempt was made to locate H atoms. However, the positions of 10 non-hydrogen-bonded H atoms, for each of the eleven PROFL molecules, were calculated with feasible stereo-chemistry and included in structure factor calculations. Hydrogen atoms were allowed to ride with fixed U_{iso} temperature factors equal to the U_{eq} of their carrier atom.

Final refinements were carried out by block full-matrix least-squares calculations assuming isotropic temperature factors for 17 of the 21 water molecules.

The refinement converged at $R = 0.121$ for 13487 selected observed data, and $R = 0.172$ for all 22852 observed data. The goodness of fit S is 2.193 (2006 parameters). Heights in final difference Fourier map are $\rho_{min} = -0.55$ e Å^{-3} and $\rho_{max} = 2.26$ e Å^{-3}.

A perspective view of the molecular structure of PROFL (Figure 3) was prepared using ORTEP-3.[27] The numbering scheme used for the proflavine rings is as used by Obendorf *et al*. (1974).[28]

The proflavine moiety is charged and the positive formal charge resides: i) on the central ring nitrogen N10; ii) part on the central ring nitrogen N10

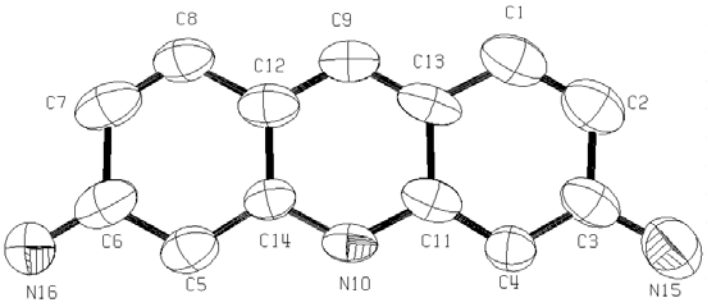

Figure 3. A perspective view of the proflavine molecule with the atomic numbering scheme used.

and part on the amino groups N15, N16. These atoms are within 3.00 Å from the nearest oxygen atom belonging to the anionic phosphate group. Because of this short distance, electrostatic interactions are likely to be as important in stabilising the structure as the considerable stacking interactions. This has the effect of the molecule having the *m* symmetry along C9–N10 or, in contrast, of making the molecule markedly asymmetric (for example, in one on the proflavine molecules C3–N15 has a length of 1.35(2) Å, and C6–N16 1.46(2) Å). This pattern is in good agreement when compared with the molecular geometry observed in the structures of proflavine hemi-sulphate hydrate and proflavine dichloride dihydrate. [28,29]

The eleven proflavine molecules within the asymmetric unit are clustered around a negatively charged wire (along *c*) formed by the phosphate anions located in the planes running at $b = 1/4$ and at $b = 3/4$, respectively (Figure 4).

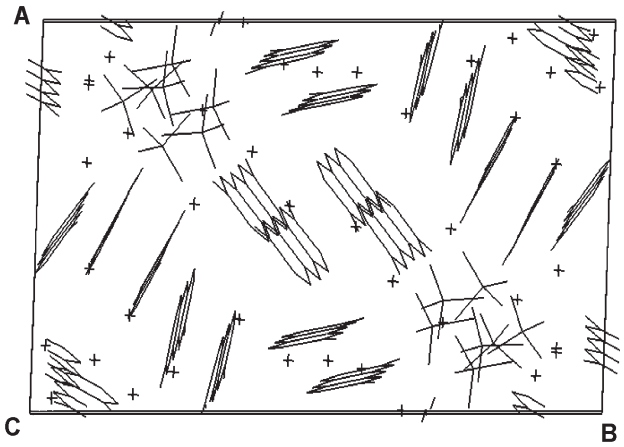

Figure 4. The *c*-axis projection of the unit cell.

The planes of the proflavine molecules are mutually inclined at roughly 50° intervals. The relative spatial disposition of the interacting proflavine molecules are: i) »off-centre parallel displaced« π-stacking with the average intermolecular distances between the centroids of 3.60–3.80 Å, and ii) »T-shaped« geometry with the average intermolecular distances between the centroids being 5.20–5.40 Å. This agrees well with a recent published analysis of aromatic amino acids in proteins and with *ab initio* and molecular mechanics calculations of benzene dimer,[30,31] showing that a »parallel displaced« structure is 0.50–0.75 Kcal/mol more stable than a »T-shaped« structure.

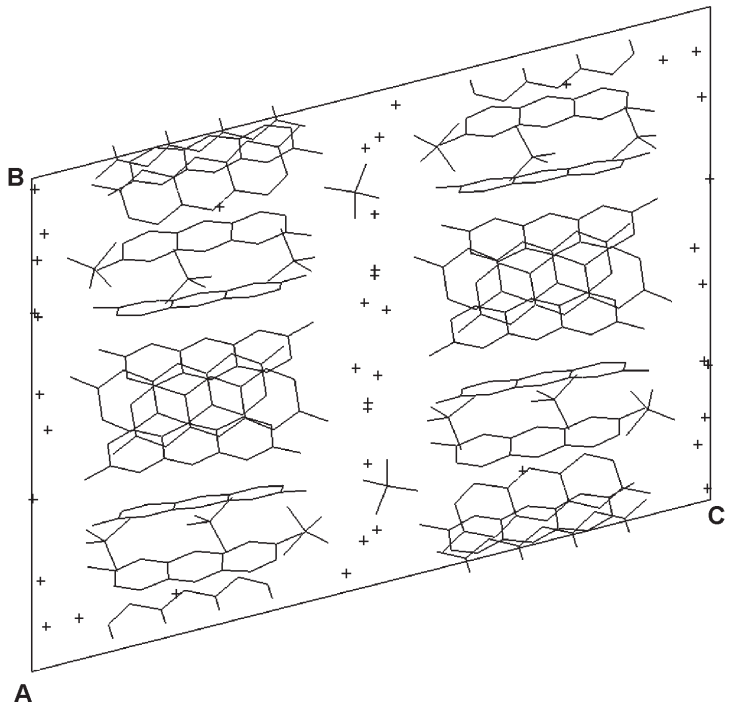

Figure 5. The *a*-axis projection of the unit cell.

Four water molecules are involved in bridging phosphate anions and proflavine molecules within the asymmetric unit and linking neighbouring symmetry related proflavine and/or water molecules. The remaining 17 water molecules excluded from these areas, having large U_{iso} values (range of 0.2–0.5 Å^2), form a complex hydrogen bonded network and are confined in the planes running at $c = 0$ and at $c = 1/2$, respectively (Figure 5).

REFERENCES

1. C. Giacovazzo, *Direct Phasing in Crystallography: Fundamentals and Applications,* Oxford University Press, 1998.
2. J. Karle and H. Hauptman, *Acta Crystallogr.* **9** (1956) 635–651.
3. G. Germain and M. M. Woolfson, *Acta Crystallogr., Sect. B* **24** (1968) 91–96.
4. C. M. Weeks and R. Miller, in: P. Bourne and K. Watenpaugh (Eds.), *Proceedings of Macromolecular Crystallography Computing School*, Bellingham, WA., 1996.
5. G. M. Sheldrick, *SHELX applications to macromolecules,* in: Suzanne Fortier (Ed.), *Direct methods for solving macromolecular structures. Proceeding of the NATO advanced study institute on direct methods for solving macromolecular structure,* Kluwer Academic Publishers, 1998, pp. 401–411.

6. M. C. Burla, M. Camalli, B. Carrozzini, G. Cascarano, C. Giacovazzo, A. G. G. Moliterni, G. Polidori, and R. Spagna, *Proceedings of the XVIIIth European Crystallographic Meeting*, Praha, Czech Republic, 16–20 August, 1998, Refs. E5-P5, p. 482.
7. C. Giacovazzo, A. Guagliardi, R. Ravelli, and D. Siliqi, *Z. Kristallogr.* **209** (1993) 136–142.
8. G. Cascarano, C. Giacovazzo, M. C. Burla, A. Nunzi, and G. Polidori, *Acta Crystallogr., Sect. A* **40** (1984) 389–394.
9. I. Glover, I. Haneef, J.-E. Pitts, S. P. Wood, D. Moss, I. J. Tickle, and T. L. Blundell, *Biopolymers* **22** (1983) 293–304.
10. Data by courtesy of R. Huber, MPI Martisried, Germany.
11. C. M. Weeks, H. A. Hauptmann, G. D. Smith, R. H. Blessing, M. M. Teeter, and R. Miller, *Acta Crystallogr., Sect. D* **51** (1995) 33–38.
12. D. A. Langs, *Science* **241** (1988) 188–191.
13. G. M. Sheldrick, Z. Dauter, K. S. Wilson, H. Hope, and L. C. Sieker, *Acta Crystallogr., Sect. D* **49** (1993) 18–23.
14. G. D. Smith, R. H. Blessing, S. E. Ealick, J. C. Fontecilla-Camps, H. A. Hauptmann, D. Housset, D. A. Langs, and R. Miller, Abstr. **E1146**, IUCr Meeting, Seattle, 1996.
15. P. J. Loll, A. E. Bevivino, B. D. Korty, and P. H. Axelsen, *J. Am. Chem. Soc*. **119** (1997) 1516–22.
16. A. Altomare, M. C. Burla, M. Camalli, G. L. Cascarano, C. Giacovazzo, A. Guagliardi, A. G. G. Moliterni, G. Polidori, and R. Spagna, *J. Appl. Crystallogr.* in press.
17. A. Albert, *The Acridines*, 2nd ed., Edward Arnold, London, 1966.
18. L. S. Lerman, *J. Mol. Biol*. **3** (1961) 18–30.
19. S. Neidle, *Prog. Med. Chem*. **16** (1979) 151–221.
20. M. Dourlent and C. Hélčne, *Eur. J. Biochem*. **23** (1971) 86–95.
21. C. Urbanke, R. Römer, and G. Maass, *Eur. J. Biochem* **33** (1973) 511–516.
22. A. R. Peacocke, in: R. M. Acheson (Ed.), *The Acridines*, John Wiley, New York, 1973.
23. S. Neidle, L. H. Pearl, and J. V. Skelly, *Biochem. J*. **243** (1987) 1–13.
24. N. C. Seeman, R. O. Day, and A. Rich, *Nature* **253** (1975) 324–326.
25. N. C. Seeman, J. L. Sussman, H. M. Bermann, and S. H. Kim, *Nature New Biol*. **233** (1971) 90–92.
26. G. M. Sheldrick, SHELX-97, Universität of Göttingen, Germany, 1997.
27. ORTEP–3, Version 1.0.2, C. K. Johson and M. N. Burnett, Windows Version 1.01b (April 1997), Dept. of Chemistry, University of Glasgow, U.K.
28. S. K. Obendorf, H. K. L. Carrell, and J. Pickworth-Glusker, *Acta Crystallogr., Sect B* **30** (1974) 1408–1411.
29. A. Jones and S. Neidle, *Acta Crystallogr., Sect B* **31** (1975) 1324–1333.
30. G. B. McGaughey, M. Gagné, and A. K. Rappé, *J. Biol. Chem*. **273** (1998) 15458–15463.
31. P. Hobza, H. L. Selzle, and E. W. Schlag, *J. Phys. Chem*. **100** (1996) 18790–18794.

SAŽETAK

Ab initio izravno određivanje faza u makromolekulskoj kristalografiji: primjena testa *Z*

Maria Cristina Burla, Carmelo Giacovazzo, Doriano Lamba, Giampiero Polidori i Giovanni Ughetto

U novije vrijeme formuliran je test *Z* kao pomoć pri prosudbi mogućnosti *ab initio* rješavanja kristalnih struktura putem direktnih metoda. Kriterij je preispitan uzimajući u obzir snažne moderne tehnike utočnjavanja faza u realnom i recipročnom prostoru. Izvještaj se odnosi na umjereno velike kristalne strukture (umjereno velike s obzirom na broj atoma) koje su nedavno riješene programom SIR97 izravnom primjenom tangentne formule.