Acta Cryst. (1997). D53, 551–557

Ab Initio Structure Determination and Refinement of a Scorpion Protein Toxin

G. David Smith,[a,b]* Robert H. Blessing,[a] Stephen E. Ealick,[c] Juan Carlos Fontecilla-Camps,[d] Herbert A. Hauptman,[a] Dominique Housset,[d] David A. Langs[a] and Russ Miller[a,e]

[a]*Hauptman-Woodward Medical Research Institute, 73 High Street, Buffalo, NY 14203, USA,* [b]*Roswell Park Cancer Institute, Elm & Carlton Street, Buffalo, NY 14263, USA,* [c]*Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853, USA,* [d]*Laboratoire de Cristallographie et Cristallogenèse des Protéines, Institut de Biologie Structurale Jean Pierre Ebel-CEA, CNRS, 41, Avenue des Martyrs, 38027 Grenoble CEDEX 1 France, and* [e]*Department of Computer Science, State University of New York at Buffalo, Buffalo, NY, 14260 USA.*

(*Received 11 February 1997; accepted 1 April 1997*)

Abstract

The structure of toxin II from the scorpion *Androctonus australis* Hector has been determined *ab initio* by direct methods using *SnB* at 0.96 Å resolution. For the purpose of this structure redetermination, undertaken as a test of the minimal function and the *SnB* program, the identity and sequence of the protein was withheld from part of the research team. A single solution obtained from 1619 random atom trials was clearly revealed by the bimodal distribution of the final value of the minimal function associated with each individual trial. Five peptide fragments were identified from a conservative analysis of the initial *E*-map, and following several refinement cycles with *X-PLOR*, a model was built of the complete structure. At the end of the *X-PLOR* refinement, the sequence was compared with the published sequence and 57 of the 64 residues had been correctly identified. Two errors in sequence resulted from side chains with similar size while the rest of the errors were a result of severe disorder or high thermal motion in the side chains. Given the amino-acid sequence, it is estimated that the initial *E*-map could have produced a model containing 99% of all main-chain and 81% of side-chain atoms. The structure refinement was completed with *PROFFT*, including the contributions of protein H atoms, and converged at a residual of 0.158 for 30609 data with $F \geq 2\sigma(F)$ in the resolution range 8.0–0.964 Å. The final model consisted of 518 non-H protein atoms (36 disordered), 407 H atoms, and 129 water molecules (43 with occupancies less than unity). This total of 647 non-H atoms represents the largest light-atom structure solved to date.

1. Introduction

Three decades ago, the number of light-atom crystal structures reported in the literature, and in *Acta Crystallographica* in particular, was quite small. The majority of light-atom structures had been solved by Patterson methods or by trial-and-error techniques, and only a relatively small number of structures, mostly centrosymmetric, had been solved by direct methods. With the advent of improved computing facilities and programs such as *MULTAN* (Germain, Main & Woolfson, 1971), the number of non-centrosymmetric light-atom structures solved by direct methods and published in the 1970's increased dramatically, and structures with as many as 40 atoms in the asymmetric unit were being solved in a fairly routine manner. Further improvements in computational techniques in the next decade extended the size of structures which could be solved in a straightforward manner to more than 100 atoms. However, the use of tangent refinement with serial phase estimation imposed limitations upon phase accuracy, and, with the exception of gramicidin A (Langs, 1988), no all-light-atom structure in excess of some 200 atoms had been solved by any method prior to 1990.

The development of the minimal function (Debaerdemaeker & Woolfson, 1983; Hauptman, 1988, 1991) and its use in the *Shake-and-Bake* method of structure determination (DeTitta, Weeks, Thuman, Miller & Hauptman, 1994) and in the *SnB* program (Miller, Gallo, Khalak & Weeks, 1994) has substantially increased the size of structures which can be solved in a relatively straightforward manner. Numerous small-molecule structures which had proved difficult to solve using conventional techniques were solved by *SnB* with no difficulty, as were two unknown peptide structures with more than 100 atoms in the asymmetric unit (Miller, DeTitta, Jones, Langs, Weeks & Hauptman, 1993) which had proved to be insoluble using other methods. Other successes include the redetermination of the gramicidin A dimer, crambin (Weeks, Hauptman, Smith, Blessing, Teeter & Miller, 1995), and the previously unknown structures, Er-1 (Anderson, Weiss & Eisenberg, 1996) and alpha-1 peptide (Privé, Ogihara, Wesson, Cascio & Eisenberg, 1995). We report here the *ab initio* solution of toxin II from the scorpion *Androctonus australis* Hector at 0.96 Å

resolution using the *SnB* program. This structure redetermination was undertaken as a test of the *SnB* program, and although the structure had been reported in the literature at a resolution of 1.8 Å (Fontecilla-Camps, Habersetzer-Rochat & Rochat, 1988) and at 1.3 Å (Housset, Habersetzer-Rochat, Astier & Fontecilla-Camps, 1994), the group in Buffalo had not been informed of the identity or sequence of the protein; the only information provided was that the structure contained approximately 500 non-H atoms and eight cysteine residues resulting in four disulfide bonds in the asymmetric unit. Given only this limited information, the *SnB* program performed remarkably well and produced an easily recognizable solution.

It is perhaps appropriate to note here that the minimal principle, upon which *SnB* is based, replaces the phase problem by one of constrained global minimization. This formulation of direct methods, as implemented in *SnB*, alternates reciprocal-space phase refinement, by reducing the value of the minimum function, with peak picking in real space (density modification) in order to impose the constraints.

2. Experimental

2.1. *Data measurement*

Toxin II (molecular weight, 7141 Da) from the scorpion *Androctonus australis* Hector crystallizes in the orthorhombic space group $P2_12_12_1$, with cell constants, $a = 45.9$, $b = 40.7$ and $c = 30.1$ Å. There are four molecules in the unit cell and the Matthews coefficient was calculated to be $1.969\ \text{Å}^3\ \text{Da}^{-1}$ (Matthews, 1968). Data were measured at room temperature from two crystals in three X-ray diffraction experiments at the DESY synchrotron in Hamburg, Germany. Both crystals were of approximate dimensions, $0.4 \times 0.4 \times 0.6$ mm, and were mounted in glass capillaries. A single set of data was measured from the first crystal on the EMBL X11 beamline ($\lambda = 0.92$ Å) with a MAR Research image plate, and two sets of data were measured from the second crystal on the EMBL X31 beamline ($\lambda = 0.72$ Å). Intensities were integrated with *MOSFLM*, Version 5.1 (Leslie, 1991) and scaled and merged with *ROTAVATA* and *AGROVATA* from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994), yielding a total of 32 890 independent reflections with an overall R_{merge} of 0.073. Completeness of data was 90.9% between 10.0 and 0.96 Å resolution and 75.6% between 0.98 and 0.96 Å resolution. Details of the data measurements are given in Table 1.

Table 1. *Data measurement and processing*

	Crystal 1	Crystal 2	
		Data set 1	Data set 2
All data			
Resolution (Å)	3.5–1.11	2.8–0.96	16.0–1.75
Total data	72171	96456	18174
$\langle I/\sigma \rangle$	8.3	6.0	4.5
Completeness (%)	91	89	96
R_{merge}	0.052	0.102	0.091
Unique data	20089	30451	5774
Data in last resolution shell			
Resolution range (Å)	1.14–1.11	0.99–0.96	1.80–1.75
$\langle I/\sigma \rangle$	4.9	1.3	3.0
Completeness (%)	81	80	87
R_{merge}	0.133	0.543	0.224

The data set supplied to the Buffalo group consisted of $|F|$ and $\sigma(|F|)$ data for 31 071 unique reflections. This represented 89.7% of the unique reflections possible to a resolution of 0.96 Å, and 92.9% of the given reflections had $|F| \geq 3\sigma(|F|)$. Normalized structure-factor magnitudes $|E|$ were obtained *via* a Wilson analysis that estimated the mean and the root-mean-square deviation from the mean of the unit-cell

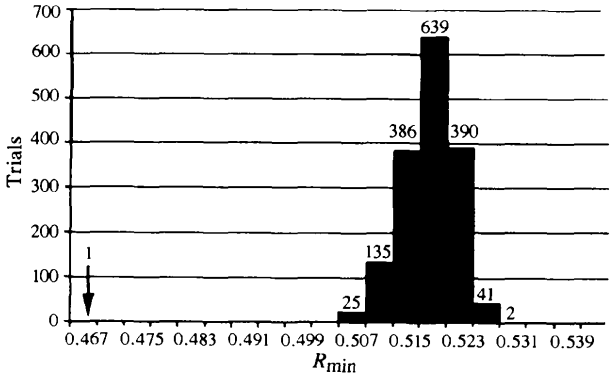

Fig. 1. Distribution of the final minimal function (R_{min}) values showing the clear separation between the solution and the non-solutions, as determined by *SnB*. The single solution is identified by the arrow at a value of R_{min} of approximately 0.467.

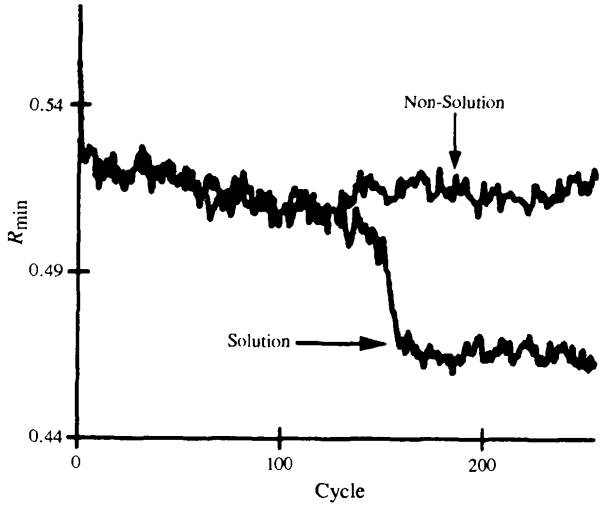

Fig. 2. Trace of the value of the minimal function (R_{min}) for the solution and a typical non-solution, as determined by *SnB*.

distribution of anisotropic atomic displacements (Blessing, Guo & Langs, 1996); the mean equivalent isotropic mean-square displacement parameter from the analysis was $\langle B_{iso} \rangle = 8\pi^2 \langle U_{iso} \rangle = 7.23\ \text{Å}^2$.

2.2. *Structure solution by SnB*

A total of 50 000 triples was generated from the largest 5 000 E magnitudes between resolution limits of ∞ and 0.964 Å. Trial structures were generated, each consisting of 200 randomly placed atoms. Each trial structure was subjected to 255 cycles of phase refinement and Fourier filtering. The Fourier filtering consisted of peak picking, while each phase-refinement cycle consisted of three passes through the phase set, utilizing a parameter-shift refinement method with a maximum of two 90° phase shifts per phase. After several weeks of processing on a network of SGI R4000 workstations, a histogram of final minimal function values (R_{min}) for every trial clearly revealed a single solution among the 1 619 trials, as shown in Fig. 1. Notice the bimodal distribution of the final minimal function values, which is typical of an *SnB* solution. This histogram clearly shows the ability of *SnB* to reach and readily identify the constrained global minimum of the minimal function. No further trials were generated following the identification of the solution. Shown in Fig. 2 is the trace of R_{min} as a function of the cycle number for the solution, as well as for a typical non-solution. The sudden drop in value of R_{min} at some point along the trace to its final and minimum value is again typical of a solution, as compared with the relatively high and constant value of R_{min} for a non-solution.

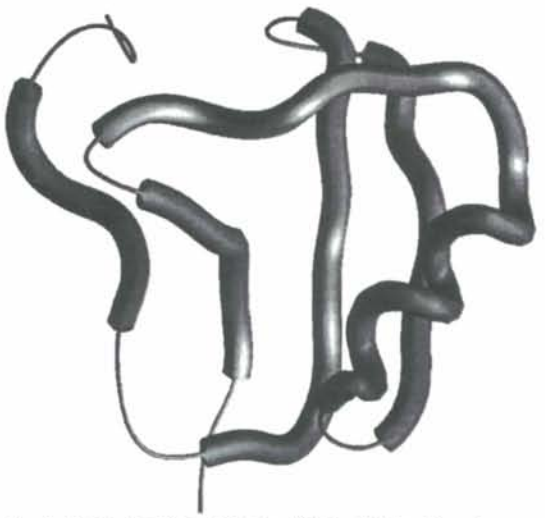

Fig. 3. A *Setor* drawing (Evans, 1993) of the five fragments, illustrated as thick tubes, obtained from the initial *SnB* solution; the unidentified portion of the protein is represented as thin tubes. The largest fragment consisted of 18 residues.

2.3. *Construction of the initial model*

A total of 500 peak positions was obtained from the initial E-map and clearly revealed many geometrical features consistent with a protein structure. In the absence of any knowledge of the amino-acid sequence or the number of residues, a very conservative approach was taken in accepting peak positions for the construction of the initial model.

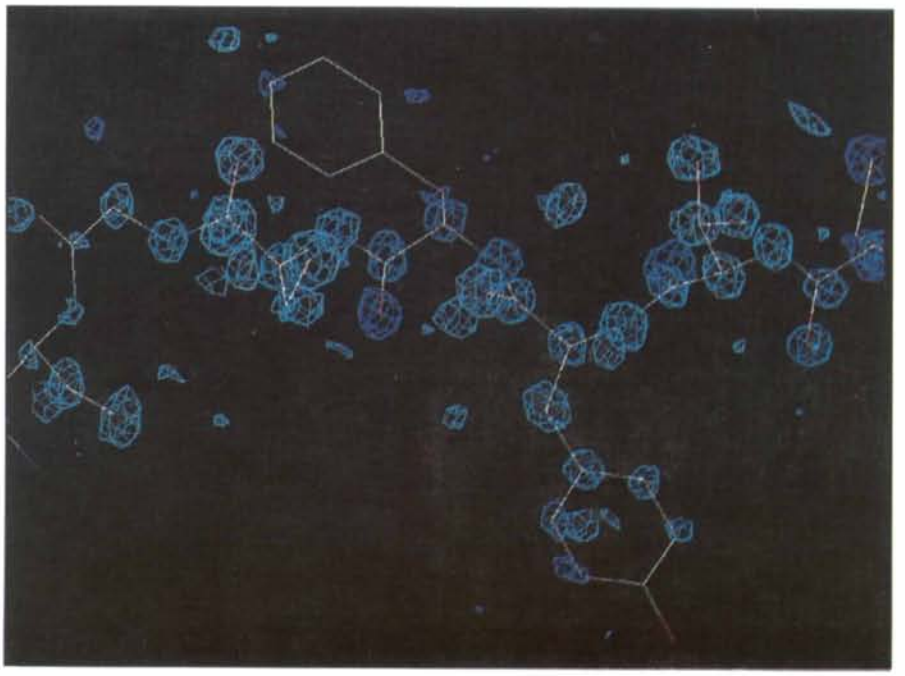

Fig. 4. A typical region of the initial E-map.

Peaks were accepted only if they were consistent with good geometry and could be used to construct an entire amino-acid residue. Five peptide fragments, shown in Fig. 3, were identified and consisted of a total of 47 residues (241 atoms) and included two disulfide bridges and one cysteine residue. One of the 47 appeared to be a sixth cysteine residue, but since there was some uncertainty regarding the second S atom of the disulfide bridge, it was modeled as glycine. The fragments varied in length from five residues to 18 residues, and 14 non-glycine residues were subsequently shown to be complete and correct. Shown in Fig. 4 is a portion of the initial *E*-map, along with the final refined structure.

2.4. *Refinement*

The 241 atoms of the 47 residues were used as the initial starting model for refinement with *X-PLOR* (Brünger, 1992*a*) and were assigned an isotropic temperature factor of 7.0 Å^2, the value obtained from the Wilson analysis (Blessing, Guo & Langs, 1996). For the *X-PLOR* refinements, 10% of the data was excluded from the refinements and used to calculate a free *R* value (Brünger, 1992*b*) for data in the resolution range of 8.0–1.0 Å; this same test set of data was used throughout all *X-PLOR* refinements. At the beginning of the refinement, the residual was 0.503 and the free *R* value was 0.514. Simulated annealing, followed by a cycle of individual-atom isotropic temperature-factor refinement reduced the residual to 0.420 and the free *R* to 0.441. The subsequent map revealed two additional residues, numerous side-chain atoms, and a third disulfide bridge for a total of 281 atoms. This procedure was repeated for nine cycles of simulated annealing and temperature-factor refinement, adding well defined protein atoms as well as water molecules. Based upon the electron-density maps, the identity of each residue was determined and the protein was found to consist of a total of 64 residues; some difficulty was encountered in identifying the fourth disulfide bridge as both cysteine residues are disordered. The final residual and free *R* were 0.203 (25 509 reflections) and 0.224 (2 830 reflections), respectively, and the model at this point contained a total of 570 atoms, of which 487 were protein atoms (24 disordered atoms) and 71 water molecules.

The resulting sequence from the refinement of the X-ray diffraction data was compared with the correct sequence and was found to contain 57 correct residues out of a total of 64, or 89%. Two of the seven sequencing errors arose from the inability to differentiate between side chains of the same size, as Asp8 had been identified as asparagine and Thr27 as valine. In a third error in sequence, the entire side chain of Lys50 was well defined with the exception of N$^\zeta$, which could not be located. The complete side chains of Asp9 and Lys50 were never located, as they are severely disordered. The side chains of Arg18 and Glu24 were also incorrectly identified as serine; subsequent refinement showed that the side chains of these two residues are disordered and have high thermal motion. It is possible that given more refinement, these two residues might have been correctly identified.

Following the correction of the sequence, several additional cycles of simulated annealing and temperature-factor refinement were performed, using a two-line weighting scheme (Smith, 1997). At the end of these rounds of refinement, the residual and *R* free were 0.198 and 0.221, respectively, for 511 protein atoms (24 disordered atoms) and 81 water molecules.

The refinement was completed using *PROFFT* (Hendrickson & Konnert, 1980; Finzel, 1987), modified to incorporate a two-line weighting scheme (GDS) and using data between 8.0 and 0.964 Å resolution; cross-validation was no longer applied. H-atom positions were calculated on the basis of geometry and were included in the refinement, and water molecules were added according to good geometrical contacts to neighboring atoms. Throughout the refinement, the geometry was continuously monitored with *PROCHECK* (Laskowski, MacArthur, Moss & Thornton, 1993). Prior to the end of the refinement, all disordered water molecules and multiple conformation side chains were excluded from the model, the structure was subjected to multiple cycles of restrained refinement, and the resulting omit maps ($2F_o - F_c$ and ΔF) were used to verify the presence or absence of all disordered atoms. The final residual was 0.158 for 30 609 data with $F_o \geq 2\sigma(F_o)$ and the weighted r.m.s. error (r.m.s. Err $= \sum w\Delta F^2/n$) in 15 equal volume shells of $\sin\theta/\lambda$ ranged from 1.15 to 1.41. A $\delta(R)$ plot (Howell & Smith, 1992) was linear with a slope and intercept of 1.245 and 0.206, respectively. Refinement statistics are given in Table 2 along with the deviations of the stereochemistry from their respective target values. As noted earlier, both main- and side-chain atoms of Cys12 and Cy63 were refined in two discrete conformations. A segment of the side chain of Asp9 was modeled in two discrete orientations, but an O atom in each conformation was never apparent from the electron-density maps. The side chain of Arg18 is poorly defined, has very high thermal motion, and is most likely disordered; attempts to refine a second orientation were unsuccessful. The side chain of Glu24, although poorly defined, was refined in two discrete conformations, both of which have high thermal motion. The N$^\zeta$ atom of Lys50 could not be located from the electron-density maps although the remainder of the side chain was well defined. The final structure consisted of 518 non-H protein atoms (36 disordered), 407 H atoms, 86 water molecules with an occupancy of unity, and 43 partially occupied water molecules in 12 sites. A ribbon drawing of the backbone of the protein is shown in Fig. 5 (Kraulis, 1991).

3. Discussion

This 64-residue protein, containing 518 non-H protein atoms and 129 water molecules (a total of 647 atoms), represents the largest light-atom structure solved to date by *ab initio* direct-methods procedures. Since the sequence and number of residues in the protein had been unknown, the 241 atoms comprising the initial model represented a very conservative starting model. But, could the initial model have been improved given the sequence of the protein? To answer this question, the peaks and the maps which were generated from the 5000 phases produced by the *SnB* program were re-examined from both objective and subjective points of view.

3.1. *Objective comparison*

Of the 500 peak positions obtained from the *SnB* program, 351 were less than 0.3 Å from a protein atom and 11 less than 0.3 Å from a water molecule. Objectively, this results in a total of 351 atom positions out of a total of 506 protein atoms (69%) that *SnB* correctly identified. Shown in Fig. 6 is a histogram illustrating the distribution of the distance between the top 500 peaks obtained from the *SnB* solution from a correctly positioned protein atom. Also illustrated is the distribution of peaks numbered from 501 to 700. The majority of the peaks are less than 0.2 Å from a protein atom, but increasing the number of peaks from 500 to 700 does not substantially increase the number of correctly placed atoms. It is possible that more atoms could have been located by increasing the number of starting phases as well as the number of triples.

Shown in Fig. 7 are main-chain atoms (N, C^{α}, C and O) of the toxin II structure which are less than 0.3 Å from the peaks obtained from the *SnB* program as well as those atoms which were not identified by *SnB*. A total of 219 out of a possible 257 main-chain atoms, or 85%, can be identified from this analysis with median, mean and r.m.s. displacements of 0.144, 0.148 and 0.163 Å,

Table 2. *Refinement statistics*

Resolution (Å)	∞–0.964
No. of reflections ($F > 0$)	31001
R	0.163
Rw	0.18
No. of reflections [$F_o \geq 2\sigma(F_o)$]	30660
R	0.162
Rw	0.186
Resolution (Å)	8.0–0.964
No. of reflections ($F > 0$)	30950
R	0.159
Rw	0.186
No. of reflections [$F_o \geq 2\sigma(F_o)$]	30609
R	0.158
Rw	0.184

	Target sigmas	Refined values
Bond distances (1–2, Å)	0.020	0.014
Angle distances (1–3, Å)	0.040	0.030
X—H bond distances (Å)	0.010	0.004
X—X—H angle distances (Å)	0.020	0.011
H—X—H (Å)	0.030	0.011
Planar distances (1–4, Å)	0.050	0.041
Chiral volume (Å^3)	0.150	0.127
Planar groups (Å)	0.020	0.022
Non-bonded distances (Å)		
Single torsion	0.500	0.192
Multiple torsion	0.500	0.201
Possible X—H$\cdots Y$ bonds	0.500	0.485
Torsion angles (°)		
Planar	3.0	4.9
Staggered	15.0	12.6
Orthonormal	20.0	19.5
Thermal parameters (Å^2)		
ΔB main chain	1.500	1.106
ΔB side chain	1.500	1.692

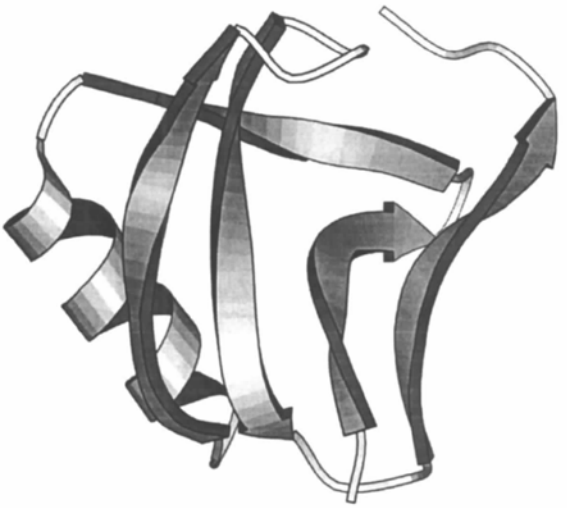

Fig. 5. A *MOLSCRIPT* drawing (Kraulis, 1991) of the refined structure.

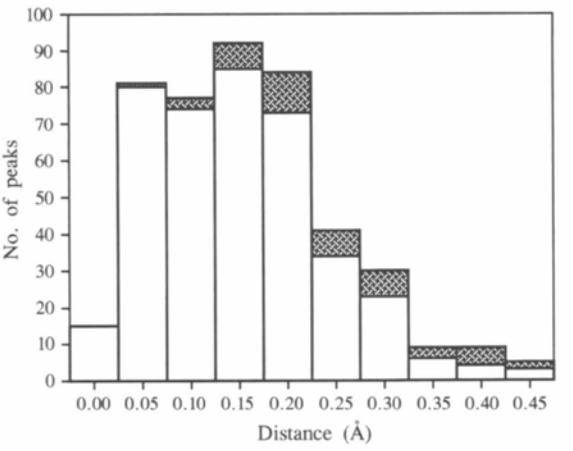

Fig. 6. Distribution of the distance between an *SnB* peak and the nearest correctly positioned protein atom. Peaks numbered from 1 to 500 are represented as white bars while those numbered from 501 to 700 are cross-hatched.

respectively. Likewise, 128 peak positions correspond to side-chain atoms and represent 52% of all side-chain atoms. If the distance of a peak to a protein atom is increased to 0.5 Å, 97% of the main-chain atoms and 69% of side-chain atoms can be identified. All S atoms could be identified, including the disordered S atoms in Cys12 and Cys63.

Finally, the mean and r.m.s. phase error between the 5000 phases determined *ab initio* by *SnB* and the phases of the refined structure were 19.1 and 29.8°, respectively. These phase errors are considerably less than those typically obtained using traditional MIR heavy-atom methods. While one might have expected to identify nearly all atoms in the protein as well as a substantial number of water molecules given the small magnitude of the phase error, this expectation was in fact not realized since only the largest 16% of the E magnitudes were used to calculate the maps whereas a much larger percentage would normally be used to generate an MIR map.

3.2. *Subjective comparison*

The above analysis was made on the basis of peak positions alone and did not utilize any structural chemical information inherent in an electron-density map. Careful examination of the maps provides a wealth of such information, and weak electron-density features not included in the top 500 or 700 peaks can be used as a guide to fit a complete side chain or to make connections between main-chain fragments. After careful examination of the maps, we estimate that 99% of all main-chain atoms and 81% of all side-chain atoms could have been accurately identified in this way, for a total of 412 atoms. Examination of Fig. 7 shows that there are numerous sites where the position of a backbone atom can be inferred from the positions of adjacent atoms and the same also is true of the side chains.

Fig. 7. Backbone atoms obtained from the *SnB* solution and less than 0.3 Å from a protein atom are illustrated as solid lines or as an x; atoms which were not identified by *SnB* are illustrated as dashed lines.

4. Conclusions

The *ab initio* structure determination and refinement of this 64-residue scorpion toxin protein clearly demonstrate the power of the minimal function as applied in the *SnB* program. In its present form, *SnB* requires data to atomic resolution, but modifications and tests are currently underway to explore its application to lower resolution data. While the automatic structure determination of this 64-residue protein by *SnB* represents a milestone in direct-methods achievements, the limitations of *SnB* are at present unknown. Future applications may well show that the minimal function as implemented in *SnB* is capable of solving larger structures at less than atomic resolution. The results of the present work represent the largest all light-atom structure solved to date, but we anticipate that this record will not stand for long.*

The authors gratefully acknowledge Dr Keith S. Wilson for his assistance at the beamlines at the EMBL in Hamburg. This research was supported by NIH grant GM-46733 and NSF grant IRI9412415.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1AHO, R1AHOSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: AM0054). At the request of the authors, the atomic coordinates will remain privileged until 8 October 1997 and the structure factors will remain privileged until 1 April 2001.

References

Anderson, D. S., Weiss, M. S. & Eisenberg, D. (1996). *Acta Cryst.* D**52**, 469–480.

Blessing, R. H., Guo, D. Y. & Langs, D. A. (1996). *Acta Cryst.* D**52**, 257–266.

Brünger, A. T. (1992*a*). *X-PLOR Manual, Version 3. A system for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT, USA.

Brünger, A. T. (1992*b*). *Nature (London)*, **355**, 472–474.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**51**, 760–763.

Debaerdemaeker, T. & Woolfson, M. M. (1983). *Acta Cryst.* A**39**, 193–196.

DeTitta, G. T., Weeks, C. M., Thuman, P., Miller, R. & Hauptman, H. A. (1994). *Acta Cryst.* A**50**, 203–210.

Evans, S. V. (1993). *J. Mol. Graph.* **6**, 244–245.

Finzel, B. C. (1987). *J. Appl. Cryst.* **20**, 53.

Fontecilla-Camps, J. C., Habersetzer-Rochat, C. & Rochat, H. (1988). *Proc. Natl Acad. Sci. USA*, **85**, 7443–7447.

Germain, G., Main, P. & Woolfson, M. M. (1971). *Acta Cryst.* A**27**, 368–376.
Hauptman, H. A. (1988). Am. Crystallogr. Assoc. Meet. Philadelphia, USA. Abstract R4.
Hauptman, H. A. (1991). *Crystallographic Computing 5: from Chemistry to Biology*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 324–332. Oxford: IUCr/Oxford University Press.
Hendrickson, W. A. & Konnert, J. H. (1980). *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 13.01–13.25. Bangalore: Indian Academy of Sciences.
Housset, D., Habersetzer-Rochat, C., Astier, J.-P. & Fontecilla-Camps, J. C. (1994). *J. Mol. Biol.* **238**, 88–103.
Howell, P. L. & Smith, G. D. (1992). *J. Appl. Cryst.* **25**, 81.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Langs, D. A. (1988). *Science*, **241**, 188–191.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leslie, A. G. W. (1991). In *Crystallographic Computing 5: from Chemistry to Biology*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 50–61. Oxford: IUCr/Oxford University Press.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M. & Hauptman, H. A. (1993). *Science*, **259**, 1430–1433.
Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
Privé, G., Ogihara, N., Wesson, L., Cascio, D. & Eisenberg, D. (1995). Am. Crystallogr. Assoc. Meet. Montreal, Québec, Canada. Abstract W008.
Smith, G. D. (1997). *Acta Cryst.* D**53**, 41–48.
Weeks, C. M., Hauptman, H. A., Smith, G. D., Blessing, R. H., Teeter, M. M. & Miller, R. (1995). *Acta Cryst.* D**51**, 33–38.