Careri, G., Fasella, P., & Gratton, E. (1975) *CRC Crit. Rev. Biochem. 3*, 141–146.
Connolly, M. L. (1981) Ph.D. Thesis, University of California, Berkeley, CA.
Connolly, M. L. (1983) *Science (Washington, D.C.) 221*, 709–713.
CRC Handbook of Biochemistry (1976) 3rd ed., p 595, CRC Press, Boca Raton, FL.
Debrunner, P., & Frauenfelder, H. (1982) *Annu. Rev. Phys. Chem. 33*, 283–299.
Frauenfelder, H., Petsko, G. A., & Tsernoglou, D. (1979) *Nature (London) 280*, 558–563.
Gurd, F. R. N., & Rothgeb, T. M. (1979) *Adv. Protein Chem. 33*, 73–165.
Hartmann, H., Parak, F., Steigemann, W., Petsko, G. A., Ponzi, D. R., & Frauenfelder, H. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4967–4971.
Hendrickson, W. A. (1985) *Methods Enzymol. 115*, 252–270.
Jolicoeur, C. (1981) *Methods Biochem. Anal. 27*, 171–287.
Karplus, M., & McCammon, J. A. (1983) *Annu. Rev. Biochem. 53*, 263–300.
Karplus, M., & McCammon, J. A. (1986) *Sci. Am. 254*, 42–51.
Kittel, C. (1971) *Introduction to Solid State Physics*, Wiley, New York.
Konnert, J. H., & Hendrickson, W. A. (1980) *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. A36*, 344–349.
Kuntz, I. D., Jr. (1975) *J. Am. Chem. Soc. 97*, 4362–4366.
Kuriyan, J. (1986) Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA.
Kuriyan, J., Wilz, S., Karplus, M., & Petsko, G. A. (1986) *J. Mol. Biol. 192*, 133–154.
Lee, B., & Richards, F. M. (1978) *J. Mol. Biol. 55*, 379.
Parak, F., Frolov, E. N., Mossbauer, R. L., & Goldanskii, V. I. (1981) *J. Mol. Biol. 145*, 825–833.
Petsko, G. A., & Ringe, D. (1984) *Annu. Rev. Biophys. Bioeng. 13*, 331–371.
Phillips, D. C. (1970) in *British Biochemistry, Past and Present* (Goodwin, T. W., Ed.) pp 11–28, Academic Press, London.
Takano, T. (1977) *J. Mol. Biol. 110*, 569–584.
Tilton, R. F., Jr., & Kuntz, I. D., Jr. (1982) *Biochemistry 21*, 6850–6857.
Tilton, R. F., Jr., Kuntz, I. D., Jr., & Petsko, G. A. (1984) *Biochemistry 23*, 2849–2857.
Walter, J., Steigemann, W., Singh, T. P., Bartunik, H., Bode, W., & Huber, R. (1982) *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. B38*, 1462–1472.

Crystal and Molecular Structure of the Serine Proteinase Inhibitor CI-2 from Barley Seeds†

C. A. McPhalen and M. N. G. James*
Medical Research Council of Canada Group in Protein Structure and Function, Department of Biochemistry, The University of Alberta, Edmonton, Alberta, Canada T6G 2H7

Received June 18, 1986; Revised Manuscript Received September 15, 1986

ABSTRACT: Chymotrypsin inhibitor 2 (CI-2), a serine proteinase inhibitor from barley seeds, has been crystallized and its three-dimensional structure determined at 2.0-Å resolution by the molecular replacement method. The structure has been refined by restrained-parameter least-squares methods to a crystallographic R factor ($=\sum||F_o| - |F_c||/\sum|F_o|$) of 0.198. CI-2 is a member of the potato inhibitor 1 family. It lacks the characteristic stabilizing disulfide bonds of most other members of serine proteinase inhibitor families. The body of CI-2 shows few conformational changes between the free inhibitor and the previously reported structure of CI-2 in complex with subtilisin Novo [McPhalen, C. A., Svendsen, I., Jonassen, I., & James, M. N. G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7242–7246]. However, the reactive site loop has some significant conformational differences between the free inhibitor and its complexed form. The residues in this segment of polypeptide exhibit relatively large thermal motion parameters and some disorder in the uncomplexed form of the inhibitor. The reactive site bond is between Met-59I and Glu-60I in the consecutive sequential numbering of CI-2 (Met-60-Glu-61 according to the alignment of Svendsen et al. [Svendsen, I., Hejgaard, J., & Chavan, J. K. (1984) *Carlsberg Res. Commun. 49*, 493–502]). The network of hydrogen bonds and electrostatic interactions stabilizing the conformation of the reactive site loop is much less extensive in the free than in the complexed inhibitor.

Serine proteinases are hydrolytic enzymes that perform a wide variety of biological functions, from activation of blood clotting to digestion of food, protein processing during viral replication, and processing of peptide hormones (Neurath, 1984). Protein inhibitors of these enzymes play important roles in regulating some of these serine proteinase activities. A knowledge of the structures of such inhibitors provides a basis for the rational design of drugs that could limit proteolysis selectively. Such small, well-characterized proteins can also provide a good system for studying protein–protein interactions and structure–function relationships through site-directed mutagenesis.

† This work was funded by grants from the Medical Research Council of Canada to the MRC Group in Protein Structure and Function at The University of Alberta. C.A.M. was the recipient of an Alberta Heritage Foundation for Medical Research Studentship. The work described in this paper formed part of the Ph.D. thesis of C.A.M. (McPhalen, 1986).

0006-2960/87/0426-0261$01.50/0

Table I: Crystal Parameters and Data Collection Conditions

	crystal 1	crystal 2
space group	*P622*	*P622*
cell dimensions		
$a = b$ (Å)	69.02 (2)	68.87 (4)
c (Å)	52.89 (1)	52.81 (3)
resolution range (Å)	21.0–2.0	59.5–2.0
total reflections measured	11362	12197
total unique reflections	5597	5903
total reflections, $I \geq \sigma(I)$[a]	3783	4785
R_{merge}[b]	0.043	0.046
diffractometer	Enraf-Nonius CAD4	
incident beam	Ni-filtered $CuK_{\bar{\alpha}}$, 40 kV, 26 mA	
diffracted beam	60-cm crystal to counter He-filled beam path	
scan type	ω scan, continuous	
scan width	0.5° at 0.67°/min	0.6° at 0.67°/min
background measurement	0.125° in ω on either side of peak scan	0.15° in ω on either side of peak scan
background correction	measured background corrected for intensity dependence and averaged over ranges of 2θ and ϕ	
max absorption correction (North et al., 1968)	1.33	1.86
decay correction	empirical function of time, max 9%	
geometric correction	Lorentz and polarization factors	
scale factor[c]	19.72	13.93
overall B[c] (Å^2)	21	21

[a] $\sigma(I) = [I + c^2I^2 + (t_I/t_{Bk})^2(\sum Bk + c^2\sum Bk^2)]^{1/2}$, where I = total intensity, Bk = background counts, t_I = time for intensity measurement, t_{Bk} = time for background measurement, and c = instrument instability constant = 0.01. [b] $R_{merge} = \sum_{hkl}(\sum_i |I_i - \langle I \rangle|/\sum_i I_i)$ for reflections measured more than once in a data set. [c] Thiessen & Levy (1973).

Chymotrypsin inhibitor 2 (CI-2)[1] is a member of the potato inhibitor 1 family of serine proteinase inhibitors according to the classification of Laskowski and Kato (1980) (Svendsen et al., 1980). It is isolated from the albumin fraction of seeds from the Hiproly strain of barley (Jonassen, 1980). The complete inhibitor consists of 83 amino acids (M_r 9250) and contains no cysteine residues (Svendsen et al., 1980). In this paper, we use the consecutive sequential residue numbering for the CI-2 molecule, and each residue number is followed by an I (for inhibitor). Thus, the reactive site bond is between Met-59I and Glu-60I. This numbering scheme differs from that given by Svendsen et al. (1984) as that alignment introduced two gaps in order to maximize homology among the several members of this family. Crystallization conditions for free CI-2 have been reported (McPhalen et al., 1983). The molecular structure of CI-2 in complex with the bacterial serine proteinase subtilisin Novo has been determined (McPhalen et al., 1985b), as has the structure of another member of the potato inhibitor 1 family, eglin-c, in complex with subtilisin Carlsberg (McPhalen et al., 1985a; McPhalen, 1986; Bode et al., 1986).

Members of five other serine proteinase inhibitor families have been studied by X-ray crystallography; structures of some of the inhibitors are known both in the free state and in complex with their cognate enzyme (Marquart et al., 1983; Hirono et al., 1984; Read et al., 1983; Bode et al., 1985). A standard mechanism of inhibition has been proposed based on detailed biochemical and kinetic data (Laskowski & Kato, 1980); the inhibitors bind to the enzymes in the manner of a good substrate ($k_{cat}/K_m = 10^4$–10^5 M^{-1} s^{-1}) but are hydrolyzed at a single peptide bond (termed the reactive site) at a very slow rate. The inhibition is a result of very tight binding ($K_{assoc} = 10^{10}$–10^{13} M^{-1}) but an extremely slow rate of hydrolysis and slow release of the cleaved inhibitor.

Comparison of the structures of the free and complexed inhibitor may help to elucidate further the reasons that these small proteins act as inhibitors rather than as good substrates. To a first approximation, the binding of inhibitor to enzyme occurs with minimal conformational change in either (Huber et al., 1974; Huber & Bode, 1978; Read et al., 1983). The major difference in the structures of most free and complexed inhibitors is the relative flexibility of the reactive site loops; in the free form, these loops exhibit large B factors whereas in the complexed form they are much more rigid. The structure of the complexed form of CI-2 (McPhalen et al., 1985b) confirms that the reactive site loop is bound firmly in the active site of subtilisin Novo. This report of the structure solution, crystallographic refinement, and molecular structure analysis of free CI-2 at 2.0-Å resolution provides additional data regarding the theory of inhibition of the serine proteinases by their protein inhibitors.

Experimental Procedures

Crystallization, Data Collection, and Structure Solution. Crystals of CI-2 were grown and characterized as described previously (McPhalen et al., 1983). The crystal data and data collection parameters of the two data sets used for the structure solution and refinement are given in Table I. Data from crystal 1 were used for the molecular replacement procedure and the first 64 cycles of refinement and data from crystal 2 for refinement cycles 65–133. It was hoped that the larger size of crystal 2 relative to crystal 1 would yield higher resolution data of better quality. In retrospect, as indicated by the refinement progress (Figure 1), this was not realized.

The method of molecular replacement (Rossmann, 1972) was used to solve the phase problem for CI-2 following a protocol that has been successful for the solution of other structures in this laboratory (Read et al., 1983). The search model was the CI-2 molecule from the crystal structure of CI-2 complexed to subtilisin Novo, at a stage at which it was partially refined to an R factor[2] of 0.193 at 2.1-Å resolution (McPhalen et al., 1985b). The rotation search was performed with the fast rotation function (Crowther, 1972; program

[1] Abbreviations: CI-2, chymotrypsin inhibitor 2 from barley seeds; rms, root mean square; B factor, thermal motion parameter = $8\pi^2\bar{U}^2$, where $\bar{U}^2$ is the mean square amplitude of vibration; I, structure factor intensity; SDS, sodium dodecyl sulfate; an I follows the sequence number of all residues in the inhibitor.

[2] $R = \sum||F_o| - |F_c||/\sum|F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively.

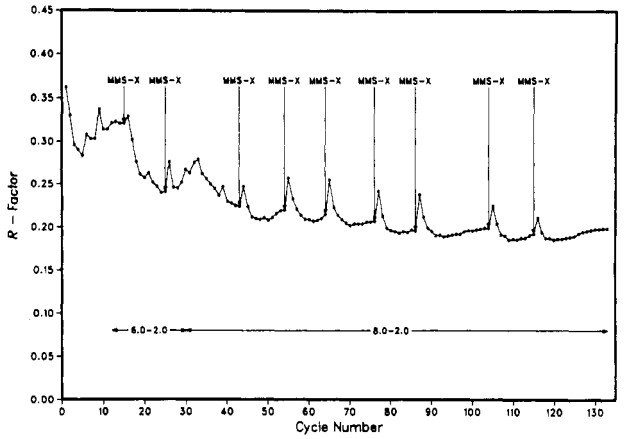

FIGURE 1: Progress of CI-2 refinement. The R factor at each least-squares cycle is plotted vs. cycle number. The resolution range of the data included at each cycle is indicated. Data from 6.0 to 2.8 Å were used for cycles 1–5, from 6.0 to 2.5 Å for cycles 6–8, and 6.0–2.2 Å for cycles 9–11. Electron density maps with coefficients $2m|F_o| - D|F_c|$, α_c (Read, 1986) were calculated and used to refit the model on the graphics system at the points marked "MMS-X".

modified by E. Dodson); the highest peak in the rotation function map was 5.6 standard deviations (σ) above the mean and 2.3 σ above the next highest peak for data in the resolution range 10–3 Å. The radius of integration was from 4 to 18 Å with the model placed in a P1 cell of dimensions 42 × 42 × 42 Å. The translation search algorithm was a correlation coefficient search on $|F|^2$ (Fujinaga & Read, 1986); the highest peak in the translation function map was 4.8 σ above the mean and 1.1 σ above the next highest peak for the 4–5-Å data, using a 1-Å sampling grid. Refinement of the rotational parameters of the solution about the translation function peak gave a final solution 5.9 σ above the mean.

The R factor for the transformed coordinates of the CI-2 molecular replacement model was 0.50 for all data to 2.0-Å resolution. This model was used in the calculation of structure factor amplitudes and phases for the computation of the initial electron density map. All electron density maps used in the structure solution and refinement were computed with coefficients designed to suppress model bias that results from phasing by partial structures with errors (Read, 1986). The MMS-X interactive graphics (Barry et al., 1976) with the macromolecular modeling system M3, developed by C. Broughton (Sielecki et al., 1982), was used for all map interpretation and model fitting.

The initial electron density map from the molecular replacement solution indicated that no large conformational changes were observed in most regions of the CI-2 molecule and only minor adjustments were made. The electron density corresponding to the reactive site loop, Thr-55I–Glu-60I, was of poor quality, and no reinterpretation could be made at this stage. The adjusted model was used to begin least-squares refinement.

Refinement. The restrained-parameter least-squares refinement program of Hendrickson and Konnert (1980), modified locally by M. Fujinaga for the FPS164 attached processor, was used. The general refinement strategy and the parameters restrained are the same as those for the refinement of the enzyme penicillopepsin (James & Sielecki, 1983). The progress of the refinement is shown in Figure 1, and the final refinement parameters and results are given in Table II.

No electron density was seen for residues 1I–18I in Fourier maps at any stage of the refinement. CI-2 is subject to multiple hydrolytic cleavages during purification, at Asn-11I, Gly-15I, and Arg-18I (Svendsen et al., 1980). The product

Table II: Final Refinement Parameters and Results

no. of cycles	133
R factor	0.198
resolution range (Å)	8.0–2.0
no. of reflections [$I \geq \sigma(I)$]	4471
no. of protein atoms	521
no. of solvent atoms	64
no. of variable parameters	2405
$\langle\|\|F_o\| - \|F_c\|\|\rangle$	40
⟨coordinate shift⟩ (Å) in final cycle	0.012
⟨B factor shift⟩ (Å^2) in final cycle	0.25
rms deviations from ideal values[a]	
distance restraints (Å)	
bond distance	0.007 (0.008)
angle distance	0.029 (0.016)
planar 1–4 distance	0.021 (0.016)
plane restraint (Å)	0.014 (0.012)
chiral-center restraint (Å^3)	0.123 (0.080)
conformational torsion angle, planar (ω) (deg)	3.0 (2.8)
nonbonded contact restraints (Å)	
single torsion contact	0.297 (0.400)
multiple torsion contact	0.276 (0.400)
possible hydrogen bond	0.240 (0.400)
isotropic thermal factor restraints (Å^2)	
main-chain bond	1.956 (2.000)
main-chain angle	2.909 (2.000)
side-chain bond	3.611 (3.000)
side-chain angle	5.285 (3.000)

[a] The values of σ, in parentheses, are the input estimated standard deviations that determine the relative weights of the corresponding restraints (Hendrickson & Konnert, 1980).

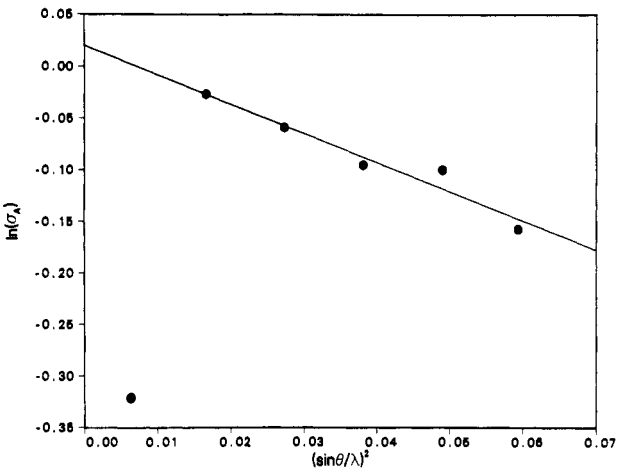

FIGURE 2: σ_A plot to estimate coordinate error. The slope of the straight-line portion of this plot is $[-\pi^3(\langle|\Delta r|\rangle)^2]$ (mean coordinate error) or $[-(8\pi^2/3)\langle|\Delta r|^2\rangle]$ (rms coordinate error) (Read, 1986). The line is determined by a least-squares fit to all data points but the first one. The first data point describes low-resolution data that are affected by omission of bulk solvent from the protein model.

with a "ragged" N-terminus was used to grow crystals. SDS–polyacrylamide gel electrophoresis of CI-2 from crushed crystals vs. CI-2 in the lyophilized powder from which crystals were grown shows that the inhibitor in the crystals has been hydrolyzed more extensively than in the powder (McPhalen, 1986). It is likely that the first 18 residues of CI-2 are not present in the crystallized inhibitor.

The refinement of CI-2 was halted at cycle 133. The model had essentially stopped changing; the indicated rms coordinate shift at cycle 133 was 0.012 Å, and examination of a difference electron density map indicated only a few minor features that could not be interpreted in terms of possible errors in the model. The refined coordinates of CI-2 will be deposited with the Brookhaven Protein Data Bank (Bernstein et al., 1977).

Quality of the Refined Structure. The accuracy of the atomic coordinates in the refined CI-2 structure was estimated

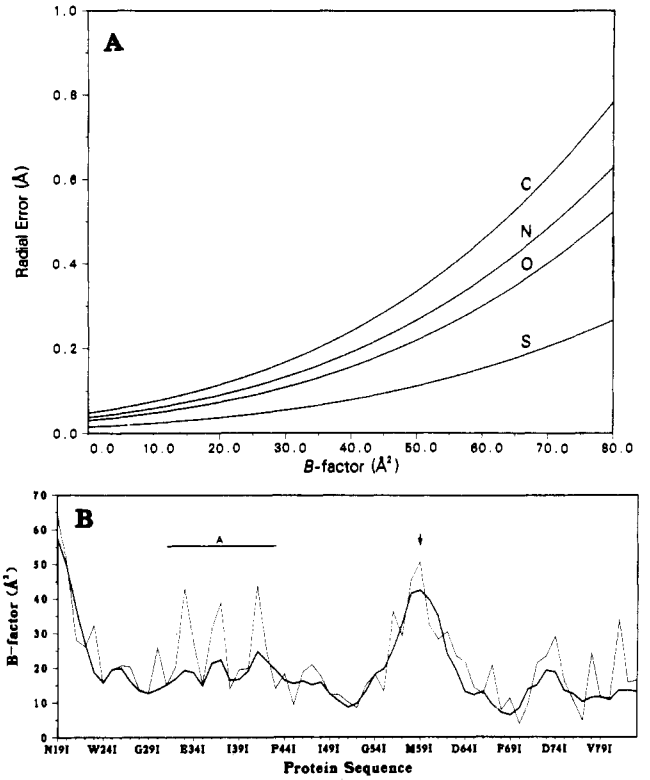

FIGURE 3: (A) Atomic coordinate error by the method of Cruickshank (1949, 1954, 1967). The estimated radial standard deviations in atomic position are given as a function of *B* factor. The four curves, from top to bottom, are for carbon, nitrogen, oxygen, and sulfur atoms in the final refined structure of CI-2. (B) Variation in *B* factor along the polypeptide chain. The heavy line denotes the mean *B* factor of the main-chain atoms and the light line that of the side-chain atoms. No *B* factors are given for residues 1I–18I of the inhibitor; they are not seen in the electron density map. The mean overall *B* factor for the atoms of the protein is 21 Å^2. The arrow denotes the reactive site bond, Met-59I–Glu-60I; the bar labeled A denotes the position of the α-helix, Ser-31I–Lys-43I.

by two methods: the σ_A plot of Read (1986) and the method of Cruickshank (1949, 1954, 1967). The σ_A plot provides an overall estimate for the errors in the coordinates of a structure, and the data for CI-2 are given in Figure 2. The mean coordinate error from this plot is 0.28 Å, corresponding to an rms error of 0.30 Å.

Individual atomic errors were estimated with the method of Cruickshank. Radial errors were calculated for each atom type in the structure, over a range of *B* factors. A plot of radial error as a function of *B* factor is given in Figure 3A; an estimate for the accuracy of the coordinates in a particular region of the CI-2 molecule can be obtained with the additional information depicted in Figure 3B, showing the variation in *B* factor along the polypeptide chain. The overall rms coordinate error for free CI-2 can be estimated as the rms value of the individual radial coordinate errors of its atoms. The value is 0.24 Å, close to that derived from the σ_A plot.

The quality of a crystal structure may also be assessed by the deviation of the geometry of the model structure from ideal values. Values for deviations from ideality of some geometrical parameters for CI-2 are given in Table II. These deviations are all small. The distribution of the main-chain torsional angles ϕ and ψ is shown in Figure 4, a Ramachandran plot (Ramakrishnan & Ramachandran, 1965). The three non-glycine residues outside the allowed conformational regions are in segments of polypeptide chain with high *B* factors and weak electron density, i.e., affected by larger coordinate errors, so their torsion angles are less well-defined.

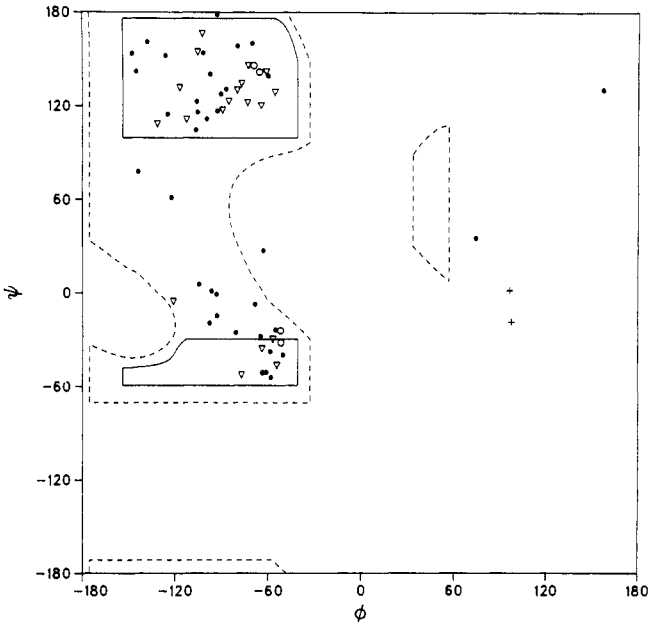

FIGURE 4: ϕ–ψ plot for CI-2. The solid lines enclose the fully allowed conformational regions for τ (N–C$^\alpha$–C) of 115°; the dotted lines enclose the more permissive regions of smaller acceptable van der Waals' contacts (Ramakrishnan & Ramachandran, 1965). The symbols denote prolines (○), β-branched amino acids (▽), glycines (+), and all other amino acids (●). Non-glycine residues outside the allowed conformational regions and their ϕ–ψ angles are Leu-20I (157, 130), Met-59I (−63, 27), and Asp-74I (74, 35).

One final indication of the quality of a crystallographic structure is the *R* factor. The *R* factor for the refined structure of the free CI-2 molecule, for all data in the resolution range 8.0–2.0 Å with $I \geq \sigma(I)$, is 0.198. This value was calculated from the 4471 reflections used in the refinement; these represent 76% of the unique set of reflections measured. The *R* factor calculated on the complete unique set of data is 0.293. The criteria of quality discussed in this section indicate that for the data available the refinement has converged and that errors in the final model are relatively small. The inability to improve the *R* factor after extensive refinement with either of the two independently measured data sets (see Table I and Figure 1) seems to be due to partial disorder for some segments of the molecule (Figure 3B), correlated with high *B* factors and weak electron density in those regions. This results in difficulty in fitting an accurate model to those segments of the electron density. A model including alternative conformations for some of the amino acids may be more appropriate for these parts of CI-2.

Results and Discussion

Secondary and Tertiary Structures of CI-2. CI-2 was the first member of the potato inhibitor 1 family to have its structure elucidated by X-ray crystallographic methods (McPhalen et al., 1985b); it is a wedge-shaped disk of approximate dimensions 28 × 27 × 19 Å, and the reactive site loop (the region containing the bond cleaved by serine proteinases) forms the narrow end of the wedge. The structure of the free CI-2 molecule shown in Figure 5 is similar in most regions to that of CI-2 from the complex with subtilisin Novo (McPhalen et al., 1985b). Table III lists the secondary structural elements of the free CI-2 molecule.

The major secondary structural features are a central 4-stranded mixed parallel and antiparallel "pseudo"-β-sheet, flanked on one side by an α-helix of 3.6 turns and on the other side by the reactive site loop in an extended conformation. The

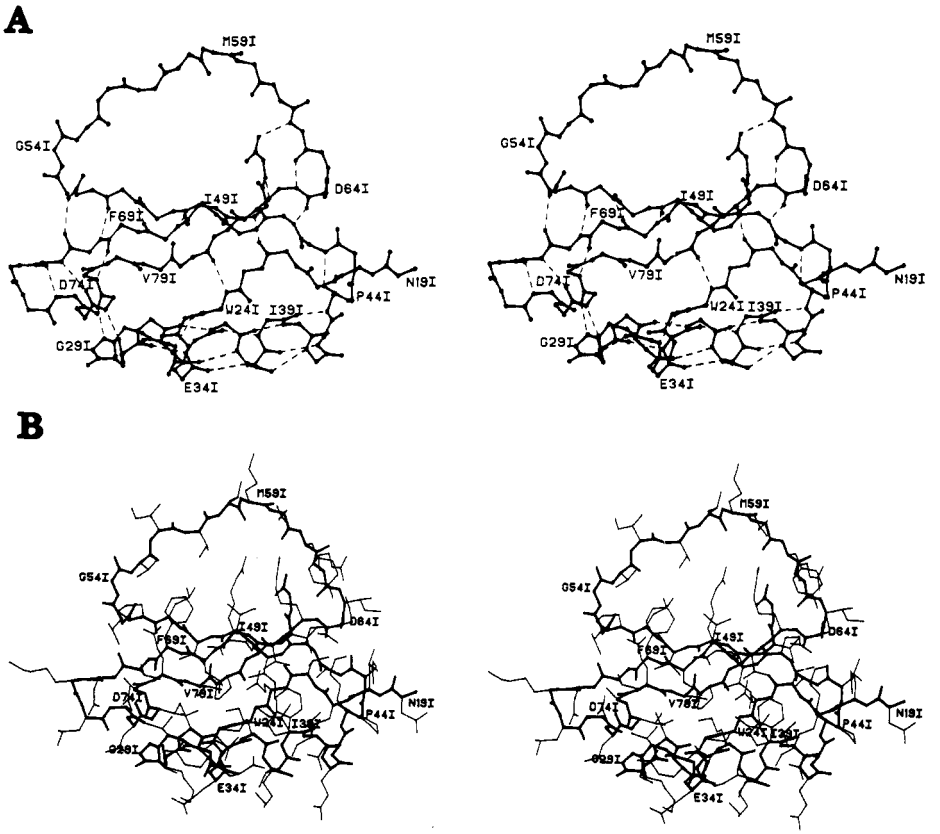

FIGURE 5: Views of CI-2. Both views include only residues seen in the electron density map of the free CI-2 molecule, Asn-19I–Gly-83I. (A) Main-chain atoms of CI-2 and their hydrogen-bonding interactions. Hydrogen bonds are drawn with dashed lines. (B) All non-hydrogen atoms of the CI-2 molecule. The main chain is drawn with heavy lines and the side chains with light lines. Every fifth amino acid residue is labeled in both views. The reactive site bond lies between Met-59I and Glu-60I.

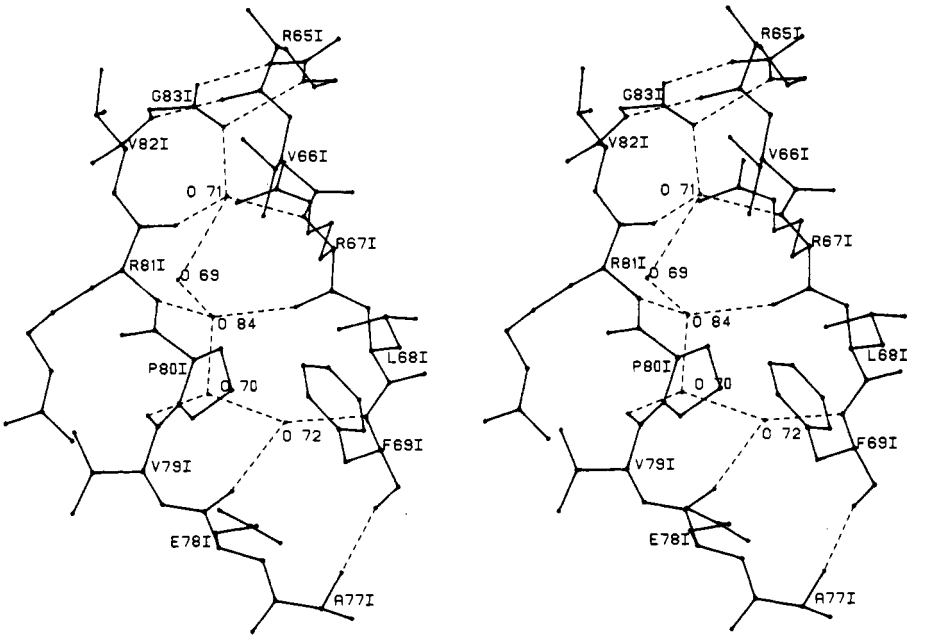

FIGURE 6: Water molecules providing hydrogen-bonding bridges (dashed lines) between two strands of the pseudo-β-sheet in free CI-2.

α-helix is not completely regular; some of the hydrogen bonds of the helix are long (Table III). The amide nitrogen atoms participating in the long bonds also make possible long hydrogen bonds to the residue i-3, in a 3_{10} turn pattern. The pseudo-β-sheet is composed of two antiparallel strands joined by a β-bridge (two hydrogen bonds in a β-sheet conformation, Kabsch & Sander, 1983) and two regular parallel strands. The two sets of strands are joined by one β-bridge and a channel of five well-ordered water molecules involved in hydrogen-bonding bridges (Figure 6). The hydrophobic core of CI-2 is located at the interface of the α-helix and the β-sheet. It is made up of 1 aromatic and 15 aliphatic residues.

The conformation of the reactive site loop of CI-2 is stabilized by a network of hydrogen bonds and electrostatic interactions, shown in Figure 7. The side chains of Arg-65I and Arg-67I extend toward the loop from the β-sheet, providing most of the contacts between the loop and the body of the inhibitor. The majority of proteinase inhibitor families

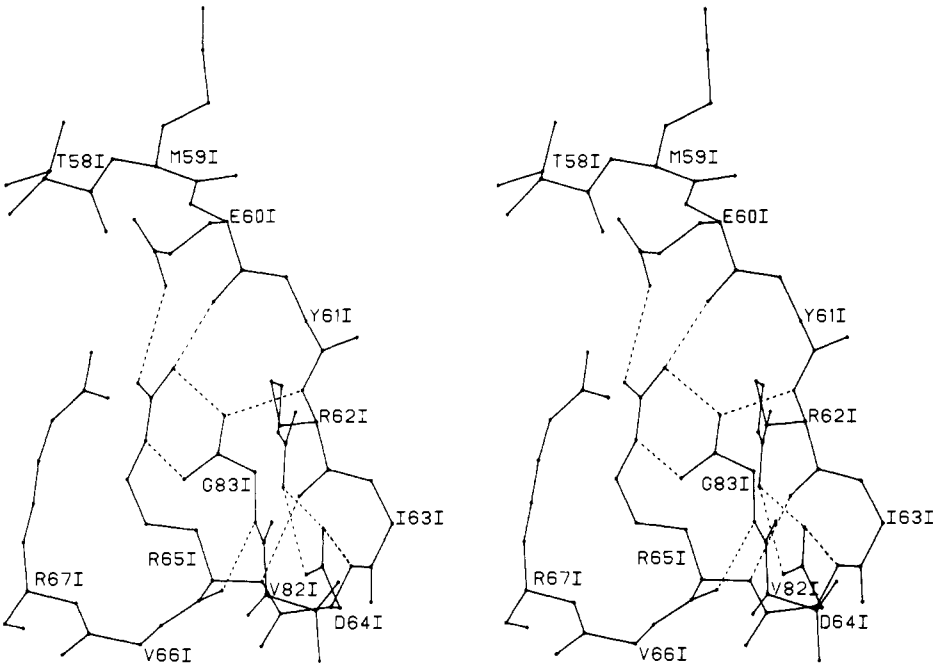

FIGURE 7: Hydrogen bonding of the reactive site loop in CI-2. Dashed lines indicate hydrogen bonds. Side chains of residues not involved in hydrogen bonding have been omitted for clarity.

Table III: Secondary Structural Elements of CI-2

element	residues[a]	N-H···O length (Å)	N-H-O angle (deg)
parallel β-sheet	Gln-47I–Asp-64I	2.62	153
	Val-66I–Gln-47I	2.84	172
	Ile-49I–Val-66I	2.89	172
	Leu-68I–Ile-49I	2.54	168
	Leu-51I–Leu-68I	2.80	165
	Val-70I–Leu-51I	2.81	165
	Val-53I–Val-70I	2.99	151
antiparallel β-bridges	Val-82I–Thr-22I	2.86	151
	Trp-24I–Pro-80I	2.91	163
	Asp-71I–Asn-75I	3.11	158
	Ile-76I–Lys-30I	2.75	158
	Ala-77I–Phe-69I	2.83	172
α-helix	Ala-35I–Ser-31I	2.80	164
	Lys-36I–Val-32I	2.90	151
	Lys-37I–Glu-33I	3.43	170
	Val-38I–Glu-34I	3.32	142
	Ile-39I–Ala-35I	3.09	154
	Leu-40I–Lys-36I	2.79	163
	Gln-41I–Lys-37I	3.44	133
	Asp-42I–Val-38I	3.10	158
	Lys-43I–Ile-39I	2.62	136
type I turns	Ala-46I–Lys-43I	2.95	162
	Arg-65I–Arg-62I	3.11	157
	Asp-74I–Asp-71I	2.88	157
type II turn	Lys-30I–Leu-27I	3.11	161
type III turn	Leu-27I–Trp-24I	2.83	151
unclassified	Gly-29I–Ile-76I	2.64	159
	Arg-62I–Gly-83I	2.78	148
	Gly-83I–Arg-65I	2.96	158

[a] The residue listed first in each pair is the hydrogen-bond donor.

have disulfide bridges flanking the reactive site bond (Laskowski & Kato, 1980). One role that these disulfides might play is to prevent the carboxylate and amino ends of a hydrolyzed reactive bond from diffusing too far to be rejoined by the enzyme. Since CI-2 and other members of the potato inhibitor family do not have disulfide bridges, a similar role may be ascribed to the two arginine side chains, Arg-65I and Arg-67I (McPhalen et al., 1985b). The electrostatic and hydrogen-bonding interactions that these side chains make with residues on either side of the scissile bond (Figure 7) could contribute to preventing the newly formed termini of a hydrolyzed reactive bond from diffusing too far. The side chains of these arginine residues are highly ordered, partly because they are buried and closely-packed, coming from the central strand of the parallel β-sheet in the molecule.

The final model of the CI-2 crystal structure contains 64 solvent molecules, all refined as oxygen atoms to represent water molecules. Calculations of the amount of solvent to be expected in the asymmetric unit of these crystals indicate that only 25–30% of the solvent has been modeled. Almost 45% of the ordered solvent molecules within 4.0 Å of the surface of CI-2 are buried in the crystal environment, i.e., by the presence of other solvent and symmetry-related inhibitor molecules (buried solvent molecules have less than 10% of their total surface area accessible, Lee & Richards, 1971). Of these 39 buried water molecules, however, only 4 are still buried in the absence of other solvent and symmetry-related molecules. These four lie between strands of the pseudo-β-sheet in CI-2 (Figure 6; O70, O71, O72, and O84) and provide some of the hydrogen-bonding bridges stabilizing the sheet. A total of 23 of the 39 buried water molecules (59%) are found at the interfaces with symmetry-related protein molecules. The remainder are buried by other solvent molecules.

Comparison with CI-2 in Complex. The free CI-2 molecule and CI-2 from the complex with subtilisin Novo can be overlapped by a least-squares superposition of equivalent α-carbon atoms, Leu-20I–Gly-83I (program of W. Bennett). The rms deviation in atomic positions is 0.47 Å for the corresponding α-carbon atoms. The largest deviations in position, between 0.5 and 1.4 Å, occur in the reactive site loop (Figure 8). The body of the inhibitor changes very little between the complexed and free inhibitor structures; the rms deviation between equivalent α-carbon atoms, excluding residues Gly-54I–Ile-63I of the reactive site loop, is 0.30 Å. The main-chain conformational angles (ϕ, ψ) for this portion of the reactive site loop in both structures are given in Table IV.

The secondary structural elements of the free and complexed CI-2 molecules are almost identical. The assignment of hydrogen bonds and secondary structural elements for CI-2 from

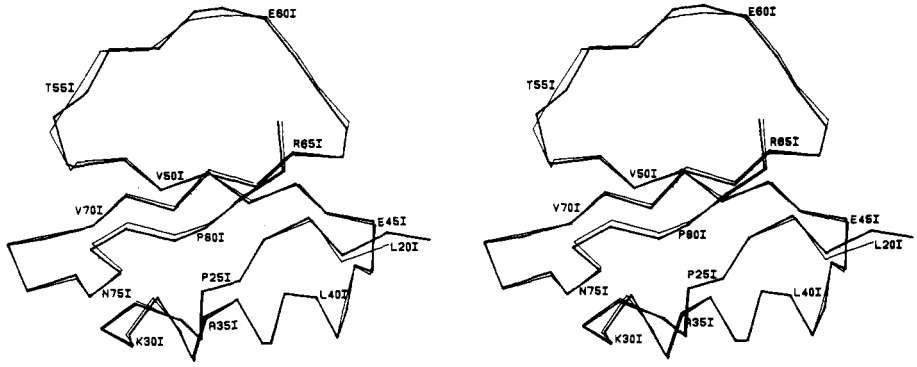

FIGURE 8: Superposition of CI-2 molecules. All equivalent α-carbon atoms of the two inhibitor molecules were overlapped as described in the text. CI-2 from the complex is drawn with thin lines and free CI-2 with thick lines.

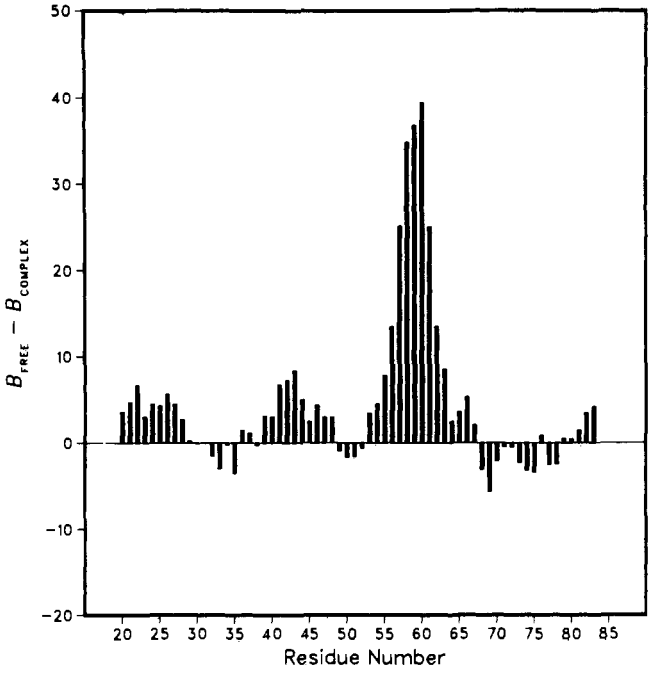

FIGURE 9: Difference *B* factor plot, free CI-2 vs. CI-2 in complex. The mean *B* factor of the main-chain atoms was calculated for each residue in free CI-2 and CI-2 from the complex with subtilisin Novo. The difference in mean *B* factor is plotted vs. residue number. No value is given for Asn-19I; no density is seen for this residue in the CI-2 structure from the complex.

the complex has changed somewhat from the initial report (McPhalen et al., 1985b), following completion of the structure refinement. In the structure of the free molecule, one hydrogen bond of a β-bridge in the pseudo-β-sheet is longer and thus weaker (Val-32I–Asp-74I, 3.45 Å), and the hydrogen bonding in the α-helix is less regular. Otherwise, even main-chain hydrogen bonds that are not part of secondary structural elements are conserved.

Some of the solvent molecules associated with the free inhibitor are conserved in the complex. If the two molecules are overlapped by the matrix relating the α-carbon atoms, there are eight water molecules in equivalent positions, within 1.0 Å of each other. Five of these are the water molecules that form the hydrogen-bonding bridges in the β-sheet (Figure 6); the other three lie close to the C-terminus of the α-helix.

Changes in the conformation of the reactive site loop between the free and complexed structures are more difficult to assess; the high *B* factors for this region in the free molecule imply large coordinate errors, on the order of half the observed rms deviations (Figure 3). The difference in the mean *B* factor of main-chain atoms for each inhibitor residue is plotted in Figure 9. The mean *B* factor for all main-chain atoms in free CI-2 is 19 Å^2; that for CI-2 from the complex is 14 Å^2. In most regions, the free CI-2 has *B* factors only marginally above or below those of the molecule from the complex. Some of the exterior loops in the free molecule are marginally better ordered than in the complex structure, probably because they are involved in contacts with symmetry-related molecules in

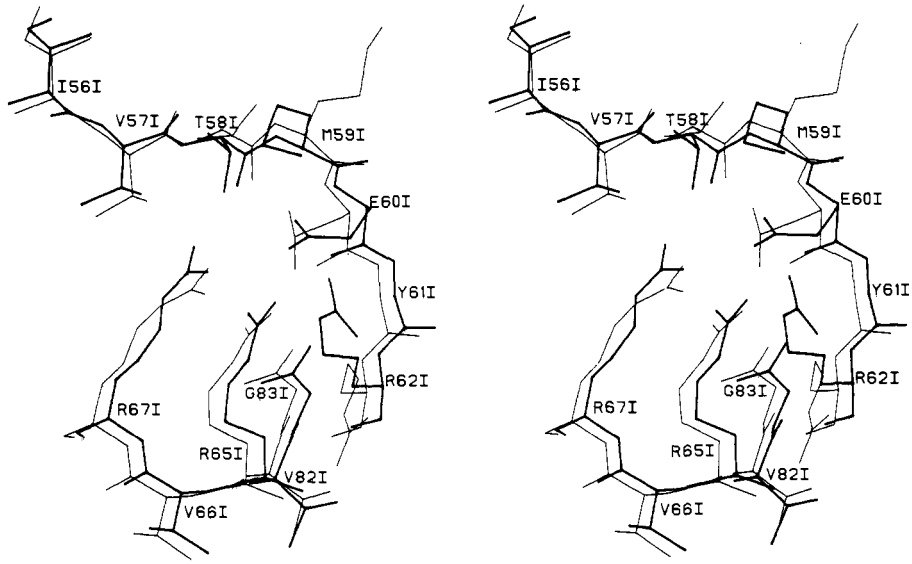

FIGURE 10: Superposition of CI-2 reactive site loops. The reactive site loops of the two inhibitors were overlapped as described in the text. CI-2 from the complex is drawn with thick lines and free CI-2 with thin lines.

Table IV: Main-Chain Conformational Angles of the Reactive Site Loop

	Gly-54I P_6	Thr-55I P_5	Ile-56I P_4	Val-57I P_3	Thr-58I P_2	Met-59I P_1	Glu-60I P_1'	Tyr-61I P_2'	Arg-62I P_3'	Ile-63I P_4'
CI-2 (free),[a] ϕ, ψ (deg)	96, 2	−80, 130	−90, 117	−103, 166	−77, 134	−63.27	−91, 128	−93, 117	−125, 115	−57, −30
CI-2 (complex),[b] ϕ, ψ (deg)	72, 30	−112, 160	−93, 140	−133, 166	−64, 147	−103, 34	−91, 146	−106, 199	−118, 113	−59, −28

[a] Present structure. [b] From the refined structure of CI-2 in complex with subtilisin Novo (McPhalen et al., 1985b; McPhalen, 1986).

the free CI-2 crystals. The reactive site loop of the free inhibitor is obviously the major region of disorder relative to the inhibitor from the complex, and the major contributor to the difference in overall *B* factor.

The differences in conformations of the reactive site loops do not disappear if a more local superposition of atoms is performed. The main-chain plus C^β atoms of residues Ile-56I–Glu-60I of the loops were overlapped; the rms deviation for this superposition was 0.51 Å. Figure 10 shows the result of the local superposition. The deviation in position of main-chain atoms is spread throughout the loop region, but the comparison of ϕ, ψ conformational angles shows that residues P_1'–P_4' have more similar conformations than those of residues P_6–P_1 (Table IV). There are several residues that have differences in ϕ, ψ of 30–40° (Thr-55I, Val-57I, and Met-59I). The conformations of several side chains in the reactive site loop are also changed between the two inhibitor structures. Thr-58I has changed by −90° in χ_1; O^{γ_1} in the free inhibitor points away from Glu-60I and forms a hydrogen bond with the side chain of Glu-78I of a neighboring molecule. Glu-60I in the free inhibitor no longer bends back toward its own main-chain NH but has a more extended conformation. Arg-62I has rotated to form a salt bridge with Asp-64I in the free inhibitor, rather than form the hydrogen bonds with Glu-60I as seen in the complex. Met-59I is in an extended conformation in the free inhibitor; the less extended conformation in the complex is presumably due to limiting contacts with the corresponding specificity pocket of subtilisin Novo.

The changes in side-chain conformations in the reactive site loop are associated with the loss, in the free inhibitor, of several hydrogen bonds from the stabilizing network in the complex, primarily those formed by Thr-58I and Glu-60I. Although it is difficult to assign cause and effect in such matters, the lack of the hydrogen bonds in the free inhibitor may allow greater movement of the reactive site loop, resulting in the high *B* factors and disorder observed in this region. The arginine residues supporting the loop, Arg-65I and Arg-67I, retain most of their hydrogen bonds between complexed and free inhibitor and are well-ordered in both structures.

The relatively high *B* factors for the reactive site loops of free inhibitors compared to the inhibitors in a complex have been observed previously for inhibitors of the Kazal family [see Read & James (1986) and references cited therein]. This indication of flexibility in the reactive site loop has been proposed as the mechanism by which inhibitors act on a wider range of enzymes; while the reactive site loop is stabilized by various hydrogen-bonding and covalent interactions to a conformation complementary to an enzyme active site, it is still able to adapt to more than one enzyme active site (Read & James, 1986). The flexibility of an inhibitor reactive site loop results in a looser binding to cognate enzymes; the entropic cost of the loss of extra internal degrees of freedom in the inhibitor reduces the available binding energy. In evolutionary terms, however, the increase in broadness of inhibitor action may be worth the decrease in tightness of binding.

ACKNOWLEDGMENTS

We thank Koto Hayakawa for growing the crystals of CI-2 and for assistance in their characterization. We also thank Randy Read, Masao Fujinaga, Anita Sielecki, and John Moult for numerous helpful discussions. Bernard Lemire prepared and ran the SDS gels discussed here.

Registry No. Chymotrypsin, 9004-07-3; proteinase inhibitor, 37205-61-1.

REFERENCES

Barry, C. D., Molnar, C. E., & Rosenberger, F. U. (1976) *Technical Memo No. 229*, Computer Systems Lab, Washington University, St. Louis, MO.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. J., Jr., Brice, M. D., Rogers, J. K., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535–542.

Bode, W., Epp, O., Huber, R., Laskowski, M., Jr., & Ardelt, W. (1985) *Eur. J. Biochem. 147*, 387–395.

Bode, W., Papamokos, E., Musil, D., Seemueller, U., & Fritz, H. (1986) *EMBO J. 5*, 813–818.

Crowther, R. A. (1972) in *The Molecular Replacement Method* (Rossmann, M. G., Ed.) pp 173–178, International Science Review 13, Gordon & Breach, New York.

Cruickshank, D. W. J. (1949) *Acta Crystallogr. 2*, 65–82.

Cruickshank, D. W. J. (1954) *Acta Crystallogr. 7*, 519.

Cruickshank, D. W. J. (1967) in *International Tables for X-ray Crystallography* (Kasper, J. S., & Lonsdale, K., Eds.) Vol. 2, pp 318–340, Kynoch Press, Birmingham, England.

Fujinaga, M., & Read, R. J. (1986) *J. Appl. Crystallogr.* (submitted for publication).

Hendrickson, W. A., & Konnert, J. H. (1980) in *Biomolecular Structure, Function, Conformation and Evolution* (Srinivasan, R., Ed.) Vol. I, pp 43–57, Pergamon Press, Oxford.

Hirono, S., Akagawa, H., Mitsui, Y., & Iitaka, Y. (1984) *J. Mol. Biol. 178*, 389–413.

Huber, R., & Bode, W. (1978) *Acc. Chem. Res. 11*, 114–122.

Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J., & Steigemann, W. (1974) *J. Mol. Biol. 89*, 73–101.

James, M. N. G., & Sielecki, A. R. (1983) *J. Mol. Biol. 163*, 299–361.

Jonassen, I. (1980) *Carlsberg Res. Commun. 45*, 47–58.

Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2577–2637.

Laskowski, M., Jr., & Kato, I. (1980) *Annu. Rev. Biochem. 49*, 593–626.

Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379–400.

Marquart, M., Walter, J., Deisenhofer, J., Bode, W., & Huber, R. (1983) *Acta Crystallogr., Sect. B: Struct. Sci. B39*, 480–490.

McPhalen, C. A. (1986) Ph.D. Thesis, The University of Alberta.

McPhalen, C. A., Evans, C., Hayakawa, K., Jonassen, I., Svendsen, I., & James, M. N. G. (1983) *J. Mol. Biol. 168*, 445–447.

McPhalen, C. A., Schnebli, H. P., & James, M. N. G. (1985a) *FEBS Lett. 188*, 55–58.

McPhalen, C. A., Svendsen, I., Jonassen, I., & James, M. N. G. (1985b) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7242–7246.
Neurath, H. (1984) *Science (Washington, D.C.) 224*, 350–357.
North, A. C. T., Phillips, D. C., & Mathews, F. S. (1968) *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. A24*, 351–359.
Ramakrishnan, C., & Ramanchandran, G. N. (1965) *Biophys. J. 5*, 909–933.
Read, R. J. (1986) *Acta Crystallogr., Sect. A: Found. Crystallogr. A42*, 140–149.
Read, R. J., & James, M. N. G. (1986) in *Proteinase Inhibitors* (Barrett, A. J., & Salvesen, J., Eds.) pp 301–336, Elsevier, Amsterdam.
Read, R. J., Fujinaga, M., Sielecki, A. R., & James, M. N. G. (1983) *Biochemistry 22*, 4420–4433.
Rossman, M. G. (1972) *The Molecular Replacement Method*, International Science Review 13, Gordon & Breach, New York.
Sielecki, A. R., James, M. N. G., & Broughton, C. G. (1982) in *Crystallographic Computing, Proceedings of the International Summer School* (Sayre, D., Ed.) pp 409–419, Carleton University, Ottawa, Oxford University Press, Oxford.
Svendsen, I., Jonassen, I., Hejgaard, J., & Boisen, S. (1980) *Carlsberg Res. Commun. 45*, 389–395.
Svendsen, I., Hejgaard, J., & Chavan, J. K. (1984) *Carlsberg Res. Commun. 49*, 493–502.
Thiessen, W. E., & Levy, H. A. (1973) *J. Appl. Crystallogr. 6*, 309.

Synthesis and Biological Properties of 5-Azido-2′-deoxyuridine 5′-Triphosphate, a Photoactive Nucleotide Suitable for Making Light-Sensitive DNA†

Robert K. Evans
Department of Microbiology/Biochemistry, The University of Wyoming, Laramie, Wyoming 82071
Boyd E. Haley*
Lucille Parker-Markey Cancer Center and Division of Medicinal Chemistry, University of Kentucky, Lexington, Kentucky 40536

Received June 24, 1986; Revised Manuscript Received September 18, 1986

ABSTRACT: A photoactive nucleotide analogue of dUTP, 5-azido-2′-deoxyuridine 5′-triphosphate (5-N_3dUTP), was synthesized from dUMP in five steps. The key reaction in the synthesis of 5-N_3dUTP is the nitration of dUMP in 98% yield in 5 min at 25 °C using an excess of nitrosonium tetrafluoroborate in anhydrous dimethylformamide. Reduction of the resulting 5-nitro compound with zinc and 20 mM HCl gave 5-aminodeoxyuridine monophosphate (5-NH_2dUMP). Diazotization of 5-NH_2dUMP with HNO_2 followed by the addition of NaN_3 to the acidic diazonium salt solution gave a photoactive nucleotide derivative in 80–90% yield. The monophosphate product was identified as 5-N_3dUMP by proton NMR, UV, IR, and chromatographic analysis as well as by the mode of synthesis and its photosensitivity. After formation of 5-N_3dUTP through a chemical coupling of pyrophosphate to 5-N_3dUMP, the triphosphate form of the nucleotide was found to support DNA synthesis by *Escherichia coli* DNA polymerase I at a rate indistinguishable from that supported by dTTP. When UMP was used as the starting compound, 5-N_3UTP was formed in an analogous fashion with similar yields and produced a photoactive nucleotide which is a substrate for *E. coli* RNA polymerase. To prepare [γ-^{32}P]-5-N_3dUTP for use as an active-site-directed photoaffinity labeling reagent, a simple method of preparing γ-^{32}P-labeled pyrimidine nucleotides was developed. [γ-^{32}P]-5-N_3dUTP is an effective photoaffinity labeling reagent for DNA polymerase I and was found to bind to the active site with a 2-fold higher affinity than dTTP. The photoactivity of 5-N_3dUMP is stable to extremes of pH, and [γ-^{32}P]-5-N_3dUTP was an effective photolabeling reagent even in the presence of 10 mM dithiothreitol. 5-Azidouracil-containing nucleotides have potential applications as active-site-directed photoaffinity labeling reagents and as tools for generating photoactive DNA and RNA to study nucleic acid binding proteins.

A variety of photoactive nucleotide analogues using aryl azides to generate nitrenes upon photolysis have been synthesized and used to study nucleotide binding proteins (Czarnecki et al., 1979; Bayley & Knowles, 1977; Guillory & Jeng, 1983). The 8-azidopurine nucleotides in particular have been used to study enzymes such as the (Na,K)-ATPase of the erythrocyte membrane (Haley & Hoffman, 1974), the Ca-ATPase of the sarcoplasmic reticulum (Briggs et al., 1980), and the beef heart mitochondrial F_1-ATPase (Wagenvoord et al., 1980). In addition, the 8-azidopurine nucleotide photoprobes have been used to label and identify amino acids in the active sites of the regulatory subunits of both type I and type II cAMP-dependent protein kinase (Bubis & Taylor, 1985; Kerlavage & Taylor, 1980) as well as in *Escherichia coli* RecA (Knight & McEntee, 1985). Because of the obvious importance of polymerases and of nucleic acid binding proteins in general, we were intrigued with the concept of developing photoactive nucleic acids to study protein–nucleic acid in-

†This work was supported by National Institutes of Health Grants GM35766 and CA 00563. This work was presented in partial fulfillment for the Ph.D. degree to R.K.E. from The University of Wyoming.

0006-2960/87/0426-0269$01.50/0