*Protein Science* (1994), *3*:1444-1463. Cambridge University Press. Printed in the USA.


# Refined structure of dimeric diphtheria toxin at 2.0 Å resolution

M.J. BENNETT, S. CHOE, AND DAVID EISENBERG

Molecular Biology Institute, Department of Chemistry and Biochemistry, and UCLA-DOE Laboratory of Structural Biology and Molecular Medicine, University of California at Los Angeles, Los Angeles, California 90024-1570

(RECEIVED March 28, 1994; ACCEPTED June 24, 1994)

**Abstract**

The refined structure of dimeric diphtheria toxin (DT) at 2.0 Å resolution, based on 37,727 unique reflections ($F > 1\sigma(F)$), yields a final $R$ factor of 19.5% with a model obeying standard geometry. The refined model consists of 523 amino acid residues, 1 molecule of the bound dinucleotide inhibitor adenylyl 3′-5′ uridine 3′ monophosphate (ApUp), and 405 well-ordered water molecules. The 2.0-Å refined model reveals that the binding motif for ApUp includes residues in the catalytic and receptor-binding domains and is different from the Rossmann dinucleotide-binding fold. ApUp is bound in part by a long loop (residues 34–52) that crosses the active site. Several residues in the active site were previously identified as NAD-binding residues. Glu 148, previously identified as playing a catalytic role in ADP-ribosylation of elongation factor 2 by DT, is about 5 Å from uracil in ApUp. The trigger for insertion of the transmembrane domain of DT into the endosomal membrane at low pH may involve 3 intradomain and 4 interdomain salt bridges that will be weakened at low pH by protonation of their acidic residues. The refined model also reveals that each molecule in dimeric DT has an "open" structure unlike most globular proteins, which we call an open monomer. Two open monomers interact by "domain swapping" to form a compact, globular dimeric DT structure. The possibility that the open monomer resembles a membrane insertion intermediate is discussed.

**Keywords:** ADP-ribosyltransferase; diphtheria; membrane insertion

Diphtheria toxin (DT) is secreted from toxic strains of the bacterium *Corynebacterium diphtheriae*. Toxigenic conversion of *C. diphtheriae* occurs by lysogenization with corynephage $\beta$ (Freeman, 1951), which carries the DT gene encoding a 535-residue protein ($M_r$ 58,342) (Uchida et al., 1971; Greenfield et al., 1983). DT kills eukaryotic cells by inactivating an essential component of the protein translation machinery, elongation factor 2 (EF-2) (Collier, 1975).

DT is produced as a protomer of 2 fragments connected by a loop containing a proteolysis site and a disulfide bond. Limited proteolysis with trypsin and reduction of the disulfide separates the 2 fragments, which are denoted fragment A and fragment B (Drazin et al., 1971). Fragment A consists of an independent folding domain, the catalytic (C) domain (residues 1–190) (Choe et al., 1992). Fragment B consists of 2 folding domains, the transmembrane (T) domain (residues 191–378) and receptor-binding (R) domain (residues 379–535) (Choe et al., 1992).

Reprint requests to: David Eisenberg, Molecular Biology Institute, University of California at Los Angeles, Los Angeles, California 90024-1570; e-mail: david@uclaue.mbi.ucla.edu.

DT has 3 functions during cell intoxication, which are performed by 1 of the 3 folding domains. First, the R domain binds to a cell-surface receptor and is endocytosed (Morris et al., 1985). The DT receptor, a 185-residue integral membrane protein, has been cloned from monkey kidney cells (Naglich et al., 1992) and has 97% sequence identity with human heparin-binding epidermal growth factor precursor (Higashiyama et al., 1992). Second, DT is triggered by low pH to undergo a conformational change and insert into the endosomal membrane, an event that has been unraveled by mimicking the process at the plasma membrane by exposing cells to low pH (Sandvig & Olsnes, 1980). In vitro, DT undergoes a cooperative conformational transition at pH 5.0, characterized by increased hydrophobicity, exposure of buried tryptophan residues, and altered susceptibility to proteases (Blewitt et al., 1985; Dumont & Richards, 1988). At low pH, DT also forms ion-conducting pores in lipid bilayers and living cells (Donovan et al., 1981; Kagan et al., 1981; Sandvig & Olsnes, 1988). Third, after being translocated to the cytosol, the C domain catalyzes the transfer of ADP-ribose from NAD to a posttranslationally modified histidine (diphthamide) of EF-2 (Van Ness et al., 1980), halting protein synthesis and killing the cell (Collier, 1975).

Although the disease diphtheria has been virtually eradicated in developed countries through vaccination, DT is still the topic of intense scientific inquiry. Knowledge about how this water-soluble protein crosses lipid bilayers may provide insight into the general principles of protein transport across biological membranes. Another active area of DT research is the production of immunotoxins, in which DT is linked to antibodies or other ligands that recognize cells targeted for destruction, such as cancer cells (Pastan et al., 1992). These areas of research depend on knowing the 3-dimensional structure of DT at atomic resolution.

## Results

### *Overall structure*

Ribbon drawings of dimeric DT and the open monomer within dimeric DT are shown in Figure 1A and B, respectively, and Kinemage 1. The structure of each of the 3 folding domains in dimeric DT is virtually identical to the corresponding domain in monomeric DT (see Kinemage 2; Bennett & Eisenberg, 1994 [companion paper]). However, the R domain of each subunit within dimeric DT is noncovalently associated with the C and T domains of the other subunit within the dimer ("domain swapping"). The crystallographic evidence for this domain-swapped structure is documented in Bennett et al. (1994) and further here, and the implications of the dimeric DT structure for DT function are discussed. A preliminary description of the structural differences between dimeric and monomeric DT has been published (Bennett et al., 1994). Here, and in the accompanying paper on monomeric DT, we describe these differences further and present many details of the structures that are relevant to substrate–NAD binding, membrane insertion, and DT–receptor binding.

The quality of the refined dimeric DT model can be assessed from the statistics given in Table 1. There is 1 half-dimer per asymmetric unit; the 2 molecules in dimeric DT are related by crystallographic 2-fold symmetry. The refined model consists of 4,021 non-hydrogen protein atoms, 43 dinucleotide (ApUp, adenylyl 3′-5′ uridine 3′ monophosphate) atoms, and 405 water molecules. Of the 535 residues in DT, our model includes 523. Twelve residues (188–199) comprising the arginine-rich proteolysis loop linking the C and T domains are not included in the model because they are disordered. The refined dimeric DT model has an $R$ factor of 19.5% for 37,727 reflections ($F > 1\sigma(F)$) between 10 and 2.0 Å resolution.

### *Accuracy of the model*

One means of estimating the accuracy of the model is to plot the $R$ factor as a function of resolution as shown for the refined dimeric DT model in Figure 2. Theoretical curves (Luzzati, 1952) suggest a mean coordinate error of 0.25 Å.

Another measure of the quality of the model is its fit to the simulated annealing (SA) omit map (Hodel et al., 1992) contoured at $1\sigma$ as shown in Figure 3 for residues 384–388. To calculate this map, residues 380–390 were omitted, as were all residues having an atom within 5.0 Å of residues 380–390. The continuous electron density for the main-chain and side-chain atoms, in addition to the well-defined density for the carbonyl oxygen atoms, show that the model is correct in this region. The

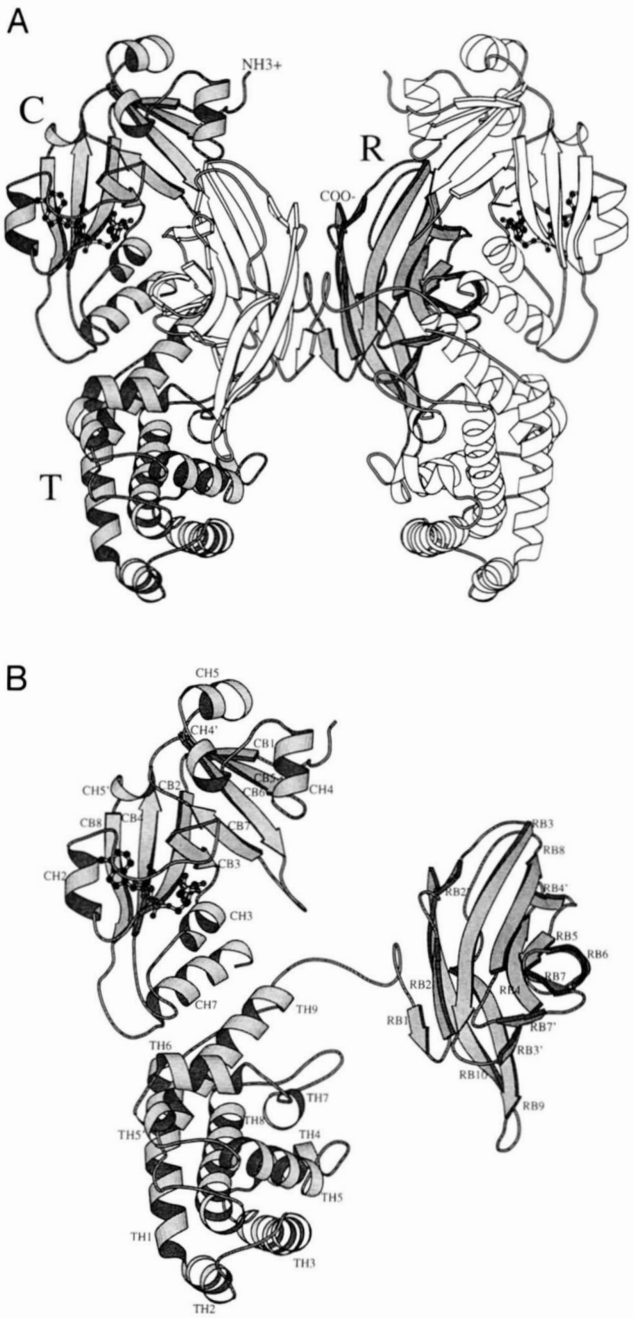


**Fig. 1.** Diphtheria toxin dimer structure. **A:** Ribbon drawing of the 2.0-Å resolution model of dimeric DT. One polypeptide chain is gray, the other is white. The termini of 1 polypeptide chain are labeled NH3+ and COO−. The domains are indicated by the letters C (catalytic), T (transmembrane), and R (receptor binding). The ApUp molecule is shown in ball-and-stick representation. **B:** Ribbon drawing of 1 monomer of dimeric DT ("open" monomer). A 3-character code used to indicate secondary structure elements refers to (1) the domain (C, catalytic; T, transmembrane; R, receptor binding), (2) the secondary structure element (H, helix; B, $\beta$-strand), and (3) the number of sequential occurrences of each type of secondary structure within each domain. Primed numbers indicate secondary structures that were not assigned in the initial 2.5-Å structure (Choe et al., 1992). The ApUp molecule is shown in ball-and-stick representation. Residues 188–199, which form a surface loop linking the carboxy-terminus of the C domain to the amino-terminus of the T domain, are disordered and not included in the model (see text). Ribbon drawings were made using MOLSCRIPT (Kraulis, 1991).

**Table 1.** *Quality of the DT dimer model refined at 2.0 Å with XPLOR*

| | |
|---|---|
| Crystallographic $R$ factor (%) | 19.5 |
| Number of reflections ($F > 1\sigma(F)$) | 37,727 |
| Resolution range (Å) | 10.0–2.0 |
| Completeness (%) | 90.4 |
| RMS deviations from target geometry | |
| Bond lengths (Å) | 0.018 |
| Bond angles (deg) | 2.6 |
| Dihedral angles (deg) | 25.8 |
| Improper angles (deg) | 2.0 |
| Number of non-hydrogen atoms | |
| Protein | 4,021 |
| Dinucleotide (ApUp) | 43 |
| Water | 405 |
| Total | 4,469 |
| Average $B$ factors (Å$^2$) | |
| All atoms | 29 |
| Protein atoms | 29 |
| Main-chain atoms | 28 |
| Dinucleotide atoms (ApUp) | 18 |
| Water molecules | 31 |

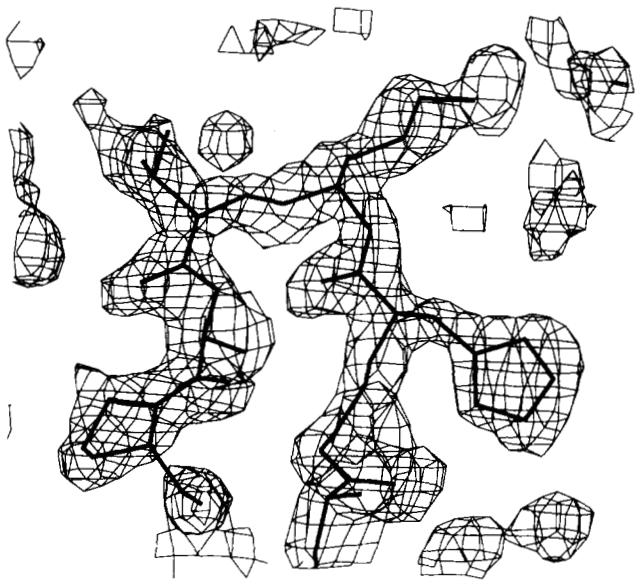

**Fig. 3.** $2F_o - F_c$ simulated annealing omit map. The SA omit map (Hodel et al., 1992) is contoured at $1\sigma$ and superimposed on the non-hydrogen atoms of residues His 384, Lys 385, Thr 386, Gln 387, and Pro 388. The omitted region included any residue with an atom within 5 Å of residues 380–390. Residues having an atom within 3 Å of the omitted region were harmonically restrained from moving into the omitted region.

map shown is representative of the quality of SA omit maps in other regions of the molecule.

### *Improvement of 3D-1D profile scores in the refined DT model*

An additional tool for model verification is a plot of the 3D-1D profile score (Bowie et al., 1991) versus residue number. The 3D-1D profile score is a measure of how compatible a residue is with its environment in the model as measured by analysis of a database of well-refined structures. A high profile score indicates that an amino acid is compatible with its environment in the model and hence is likely to be correct. Figure 4 shows a comparison of the 3D-1D profile scores of the previously published dimeric DT model determined at 2.5 Å by multiple isomorphous replacement (Choe et al., 1992) and the refined 2.0-Å dimeric DT model. The labeled minima in Figure 4 indicate several areas where the profile scores improved after refinement of the model.

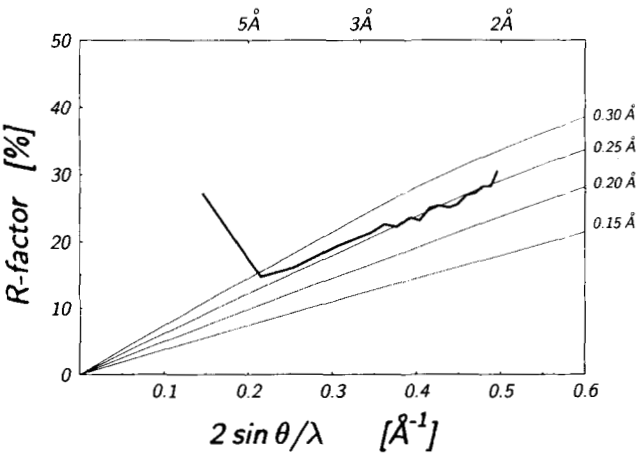


**Fig. 2.** Luzzati (1952) plot indicating a mean coordinate error of 0.25 Å for the dimeric DT model. The thick line shows $R$ factor as a function of resolution. Theoretical curves corresponding to mean coordinate errors of 0.15, 0.20, 0.25, and 0.30 Å are plotted in thin lines.

The replacement of previously misregistered segments of the model with the correct segments greatly increased the profile scores of both segments in the refined model. In the 2.5-Å model, residues 43–46, 130–135, 173–175, 238–254, 263–268, 387–390, 413–453, and 514–522 had the DT sequence misregistered in the electron density. In the refined model, they are replaced by residues 44–47, 133–138, 172–174, 239–255, 262–267, 388–391, 412–452, and 508–516, respectively. Loops flanking the residues, which replace misregistered segments (residues 43, 48, 126–132, 139–143, 167–171, 175, 232–238, 268–269, 387, 392–393, 411, 453, 496–507, and 517–523), appear as peaks or gaps in a plot of the RMS differences between the 2.5-Å model (Choe et al., 1992) and the refined 2.0-Å model after a structural alignment of the models using the program ALIGN (Satow et al., 1986) (Fig. 5). Because of the structural alignment, residues which themselves replace misregistered segments do not appear as peaks in Figure 5.

Segments where the main chain was rebuilt without shifting the sequence registration are: 7–9, 68–73, 223, 293–296, 351–354, 400–405, and 464–466. These residues also appear as peaks in Figure 5.

Excluding residues 387–393 (discussed below), all of the previously misregistered segments have higher profile scores in the refined model than the same segments in the initial model at 2.5 Å resolution, indicating that they are compatible with their

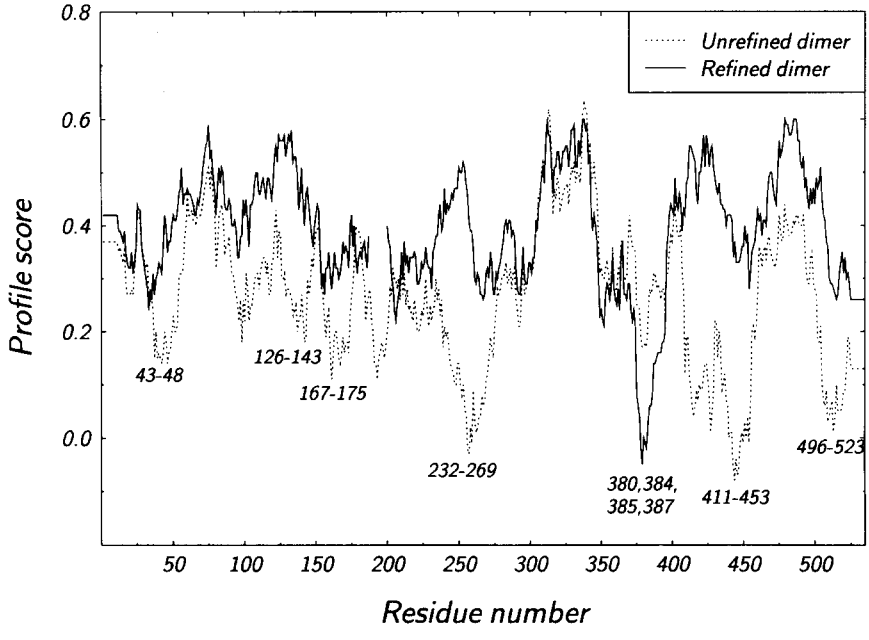


**Fig. 4.** 3D-1D profile window plot: refined versus unrefined dimeric DT. The 3D-1D profile score (Bowie et al., 1991; Zhang & Eisenberg, 1994) is averaged over a 21-residue window and plotted versus residue number for the 2.0-Å refined dimeric DT model (solid lines) and the 2.5-Å initial dimeric DT model (Choe et al., 1992) (broken lines). Residues 188–199 are not included in the refined model. Several segments are labeled for identification.

environments in the refined model (Fig. 4). For example, in the refined model, the previously misregistered residues 514–522 are replaced by 508–516, and residues 496–507 and 523 in the flanking loops are rebuilt. As shown in Figure 4, the profile scores of residues 496–523 increased in the refined model.

Profile scores of other segments increased in the refined model because residue environments became more compatible with the amino acid sequence as a result of shifting the sequence registration of a spatially contiguous segment. For example, although residues 455–495 are not themselves shifted in sequence registration relative to the initial model, their profile scores increased because of shifting residues 412–452 in the refined model, which form tertiary contacts.

The pronounced dip in the 3D-1D profile plot occurs because of the low profile scores of residues 380, 384, 385, and 387. These residues are in or near the hinge loop (residues 379–386), which changes conformation when DT dimerizes by domain swapping (Bennett et al., 1994); residues in the hinge loop have different environments in dimeric and monomeric DT. For example, Tyr 380 and His 384 are buried in monomeric DT, but are more exposed in dimeric DT, which is a less favorable environment for both Tyr and His. In addition, Lys 385 has unusual

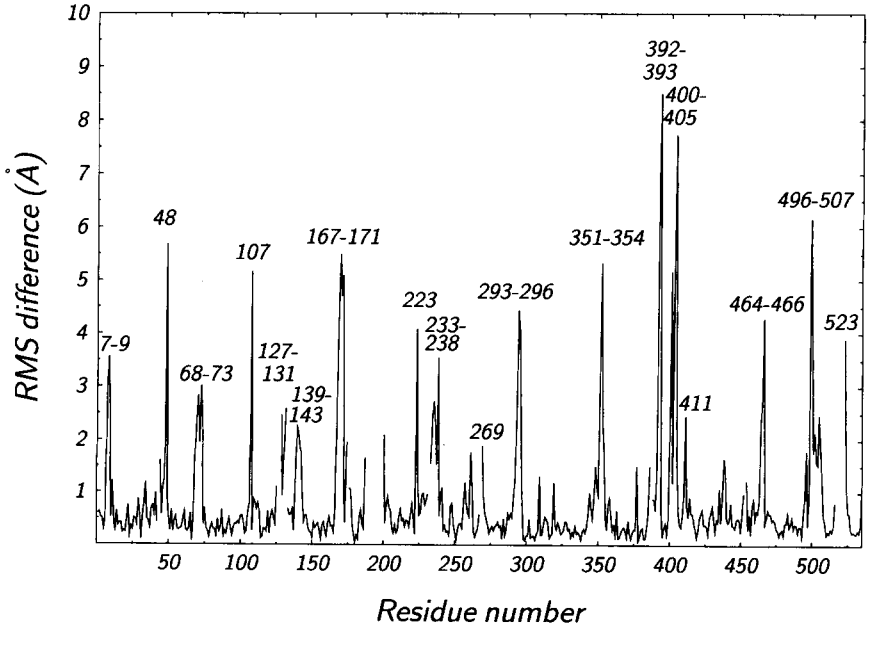


**Fig. 5.** RMS positional differences between refined and unrefined dimeric DT. The RMS differences between corresponding Cα atoms of the superimposed 2.0-Å refined dimeric DT model and the initial 2.5-Å model (Choe et al., 1992) were calculated using the program ALIGN (Satow et al., 1986), which superimposes 2 sets of Cα coordinates using a structural alignment. The overall RMS difference is 0.6 Å based on 442 Cα pairs. The comparison using ALIGN was done for the coordinate sets of 2 segments (residues 1–187 and 200–393; and 394–535); the breakdown into 2 segments was necessary because the position of the R domain relative to the C and T domains is different in the initial and refined models due to the fact that only the refined model has domain swapping. Gaps in the plot are residues that do not pair with a residue in the other model. With the exception of residues 188–199, which are deleted in the refined model, all residues plotted as gaps are in loops that changed in length during refinement to adjust the sequence registration of an adjacent secondary structure (residues 43, 126, 132, 175, 232, 268, 387, 453, 417–522). Several peaks discussed in the text are labeled for identification.

main-chain torsion angles in dimeric DT (discussed below), but not in monomeric DT. The profile scores of the hinge loop in monomeric DT are higher, with an average of approximately 0.2 (Bennett & Eisenberg, 1994). In dimeric DT, the hinge loop is extremely well defined in the $2F_o - F_c$ SA omit map (Fig. 3) and the position of the side chain of His 384 is confirmed by the binding of a mercury atom at the His N$\epsilon$ atom (Bennett et al., 1994). The low profile score of this well-defined portion of the model is consistent with the higher energy of domain-swapped dimeric DT relative to monomeric DT (Carroll et al., 1986a).

### *Main-chain conformation*

Scatter plots of the main-chain torsion angles $\phi$ and $\psi$ for both glycine and non-glycine residues are shown in Figure 6. Values for the backbone dihedral angles of the refined model cluster into the allowed regions of the Ramachandran plot, showing that the model is stereochemically sound. Morris et al. (1992) have defined "core," "allowed," and "generously allowed" regions of the Ramachandran plot based on 462 structures in the 1990 release of the Brookhaven Protein Data Bank. Figure 6 has the boundaries drawn for the allowed regions. By comparison with 119 well-refined structures at ≤2.0 Å resolution, structures refined at the same resolution are expected to have >90% of the non-glycine main-chain torsion angles in the core regions for $\alpha$-helix, $\beta$-strand, and left-handed helix, which are smaller subsets of the allowed regions shown in Figure 6. The refined DT structure has 92% of its non-glycine residues in the core region, as compared to only 65% of the non-glycine residues in the initial 2.5-Å model.

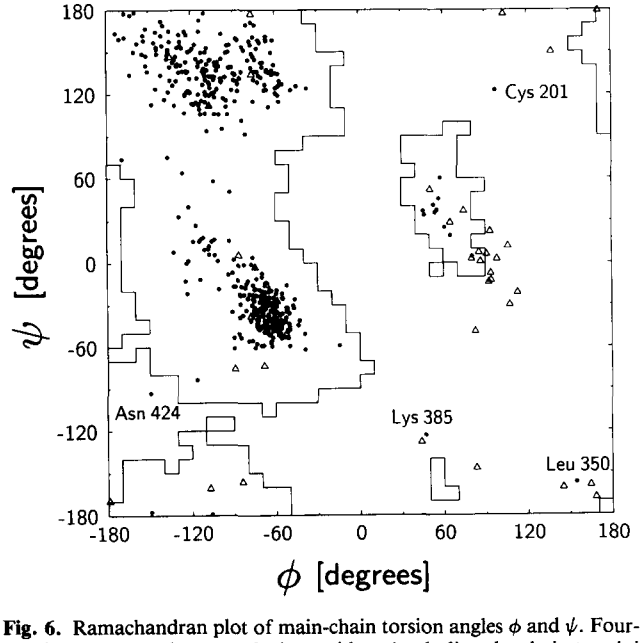


**Fig. 6.** Ramachandran plot of main-chain torsion angles $\phi$ and $\psi$. Four-hundred seventy-nine non-glycine residues (excluding the chain termini Ala 187, Ser 200, and Ser 535) are indicated with a circle and 40 glycine residues (excluding the chain terminus Gly 1) are indicated with a triangle. The allowed regions defined by Morris et al. (1992) are indicated by solid boundaries. Non-glycines outside the allowed regions are labeled: Cys 201 (97°, 123°), Leu 350 (154°, −157°), Lys 385 (47°, −123°), and Asn 424 (−150°, −93°). Seventeen of 40 glycines are in regions not allowed for residues with a side chain.

Residues that are outside the allowed regions are labeled in Figure 6: Cys 201 adopts an unfavorable conformation in order to make the disulfide bond with Cys 186; Asn 424 forms a hydrogen bonded crystal packing contact; Leu 350 is in a poorly defined loop having high main-chain *B* factors; and Lys 385 is in the hinge loop that changes conformation when DT dimerizes by domain swapping.

Ten of 479 non-glycine residues cluster in the left-handed $\alpha$-helical region near (60°, 40°). They are: Asn 16, Gln 43, Asp 129, Asn 235, Asp 318, Asp 392, Asn 453, Asp 467, Asp 519, and His 520. Five of these residues, Gln 43, Asp 129, Asn 318, Asp 392, and Asn 453, are found in the $i + 1$ position of type I′ (inverse common) turns (Richardson, 1981). Asn 16 occurs as a tight connection between the first 2 $\beta$-strands in the C domain. The remaining 4 residues are in parts of the structure where main-chain atoms are poorly defined in the electron density and have high *B* factors.

Of the glycine residues, 17 of 40 are found in regions not allowed for residues with side chains. None of the 21 prolines has a *cis* peptide bond. Proline residues are expected to have $\phi$ near −60° due to geometric constraints of the pyrrolidine ring. As expected, all prolines in DT are found either in the polyproline region at (−60°, 140°) or in the $\alpha$-helical region (−60°, −40°) of the Ramachandran plot.

Stereo figures of the C$\alpha$ backbones of each domain are shown in Figure 7. The secondary structure of the model is given in Figure 8, which lists DSSP (Kabsch & Sander, 1983) secondary structure assignments for the initial 2.5-Å model (Choe et al., 1992) as well as DSSP and visual secondary structure assignments for the refined 2.0-Å model. Hydrogen bonded reverse turns are also listed in Figure 8.

As shown in Figure 8, the DSSP assignment of residues to secondary structure classes differs between the initial 2.5-Å model (Choe et al., 1992) and the refined 2.0-Å model. These differences can be explained by 2 changes in the refined model. First, the total number of residues in secondary structures is higher in the refined 2.0-Å model than in the initial 2.5-Å model: DSSP assigns 73% of the residues in the refined model to $\beta$-strands, helices, or hydrogen bonded turns, as compared to 61% of the residues in the initial model. As an example of this increase, consider the 4 consecutive residues in the initial model (residues 20–23) forming a $\beta$-strand (Fig. 8, second line); in the refined model, 6 consecutive residues (residues 18–23) form an extended version of the same $\beta$-strand (Fig. 8, third line). The second change involves residues that replace previously misregistered segments, resulting in shifts in secondary structure boundaries. For example, the second to last $\beta$-strand in the R domain was comprised of residues 514–522 in the initial model, whereas it is comprised of residues 508–516 in the refined model. Other residues that were previously misregistered in secondary structures in the initial model are: 130–135, 238–254, 263–268, and 413–453.

Figure 8 also shows underlined residues that are assigned to named secondary structure elements. We have made additions and deletions to the previously named secondary structures in the model (Choe et al., 1992) to maximize the number of residues in named secondary structures that have appropriate DSSP assignments and to remove residues without DSSP assignments. In the C domain: the first helix assigned in the 2.5-Å model (CH1, residues 2–7) is deleted because it contains too few residues in helical conformation and residues 2–3 are instead assigned to a hydrogen bonded turn; an $\alpha$-helix (CH4′) is assigned

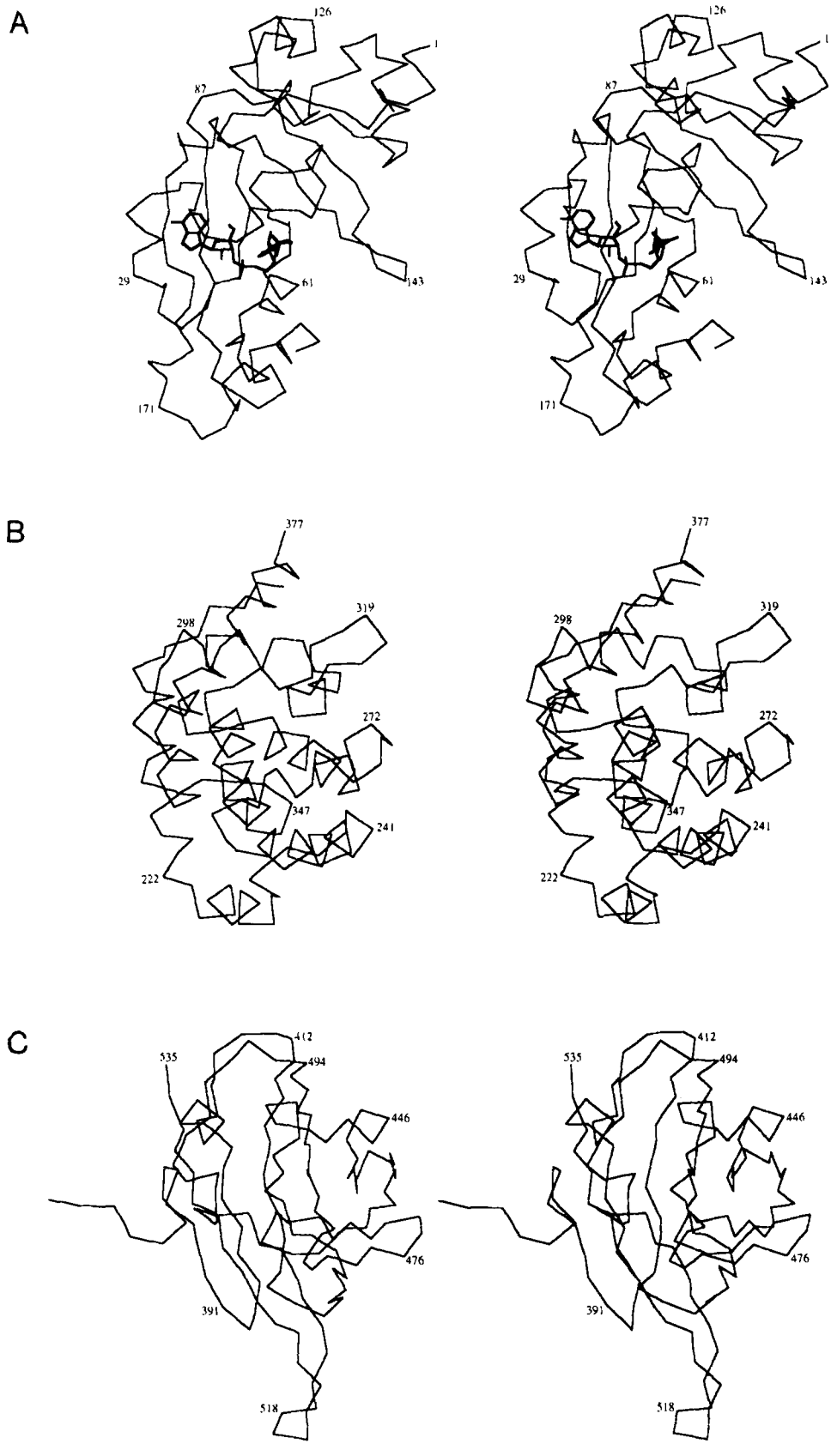


**Fig. 7.** Stereo figures of the Cα backbones of the 3 domains of DT. **A:** Catalytic domain (residues 1–187). All non-hydrogen atoms of ApUp are also shown. **B:** Transmembrane domain (residues 200–377). **C:** Receptor-binding domain (residues 378–535). All stereo figures are in approximately the same orientation as Figure 1. Some residues are labeled for identification.

```
1                GADDVVDSSK SFVMENFSSY HGTKPGYVDS IQKGIQKPKS GTQGNYDDDW
 2.5 Å DSSP      TTTTT   S    EES   E EEE TTHHHH HHH      SS S S  TTTT
 2.0 Å DSSP      GGGTB GGG  EEEET EEE EEE TT TTG GGG      SS S TT SSGGG
 2.0 Å visual    TT    TT   EEEE EEEE EEE TT  GG GGG          TT   TT
                 I     I     CB1    CB2    II    CH2             I'   I

51               KGFYSTDNKY DAAGYSVDNE NPLSGKAGGV VKVTYPGLTK VLALKVDNAE
 2.5 Å DSSP        EEEES HH HHGGG B SS  SSSS B EE EEEE SSEEE EEEES   HH
 2.0 Å DSSP      SEEEES HH HHGGG B TT STTTS B EE EEEE SSEEE EEEBS   HH
 2.0 Å visual      EEEE  HH HHGGGG  TT          EE EEEEE  EEE EEEE    HH
                   CB3    CH3       I              CB4       CB5

101              TIKKELGLSL TEPLMEQVGT EEFIKRFGDG ASRVVLSLPF AEGSSSVEYI
 2.5 Å DSSP      HHHHHSS  S SS HHHHTTS HHHHTTTTS  EEEEES S S SSS    EEE
 2.0 Å DSSP      HHHHHTT  S SS HHHHHTS HHHHHHHSTT  SEEEEEEE   TT   EEEE
 2.0 Å visual    HHHHHH         HHHHHG  HHHHHHHHTT   EEEEEEE   TT   EEEE
                   CH4            CH4'      CH5  I'     CB6    II    CB7

151              NNWEQAKALS VELEINFETR GKRGQDAMYE YMAQACAGNR VRRSVGSSLS
 2.5 Å DSSP      E HHHHTT E EEEEE SHHH HTTTT THHH HHTTSSS  S SSS   SSS
 2.0 Å DSSP      E TTGGGG E EEEEEESGGG   STTHHHHH HHGGGG *** *********
 2.0 Å visual    E TTGGGG E EEEEEE  TT    TTHHHHH HHGGGG *** *********
                   I CH5'     CB8    I     II'    CH7

201              CINLDWDVIR DKTKTKIESL KEHGPIKNKM SESPNKTVSE EKAKQYLEEF
 2.5 Å DSSP           HHHH HHHHHHHHHH SSSSTHHHHH  S SS   HH HHHHHHHHHH
 2.0 Å DSSP           HHHH HHHHHHHHHH HH HHHHHHH HHS  S   H HHHHHHHHHH
 2.0 Å visual         HHHH HHHHHHHHHH HHHHHHHHH HH        H HHHHHHHHHH
                        TH1              TH2              TH3

251              HQTALEHPEL SELKTVTGTN PVFAGANYAA WAVNVAQVID SETADNLEKT
 2.5 Å DSSP      HHHTTS TTT TTHHHHHT S GGGSHHHHHH HHHHHHTT   SS TT HHHH
 2.0 Å DSSP      HHHHTTSGGG HHHHHHTSS  GGGSHHHHHH HHHHHHTT   HHHHT HHHH
 2.0 Å visual    HHHHTT TTH HHHHHHHG   TT  HHHHHH HHHHHHHG   HHHHH HHHH
                   I  I     TH4      I        TH5          TH5' TH6

301              TAALSILPGI GSVMGIADGA VHHNTEEIVA QSIALSSLMV AQAIPLVGEL
 2.5 Å DSSP      HHHHTT S T HHHHTEETTE E   SHHHHH HHHHHHHHHH HHHHHSSS
 2.0 Å DSSP      HHHHTT S H HHHHTEETTE E   SHHHHH HHHHHHHHHH HHHHHHH  S
 2.0 Å visual    HHHHTT   H HHHH   TT      HHHHH HHHHHHHHHH HHHHHHH
                   I       TH7    I'            TH8

351              VDIGFAAYNF VESIINLFQV VHNSYNRPAY SPGHKTQPFL HDGYAVSWNT
 2.5 Å DSSP      SS TTTTTHH HHHHHHHHHH HHTTTSS     TTSSS   T TS  EEEEEE
 2.0 Å DSSP       HHHHHHTHH HHHHHHHHHH HHHHHHS     TTSTT  EE ETTEEEEESS
 2.0 Å visual          HHH HHHHHHHHHH HHHHHH      TT TT  EE ETTEEEEE
                         TH9                      II II'   RB1 I' RB2

401              VEDSIIRTGF QGESGHDIKI TAENTPLPIA GVLLPTIPGK LDVNKSKTHI
 2.5 Å DSSP        SSSSB S  TT EEEEEE  EE SSS  B   EEE B  TT TSEE TTT E
 2.0 Å DSSP      TGGGEEE SS SEEEEEEEEE EESSS EEE  EEEEE BTTT BEE TTT EE
 2.0 Å visual     TT EEEE    EEEEEEEEE EE    EEE  EEEEE       EEE    EE
                  I  RB2'         RB3        RB3'  RB4         RB4'   RB5

451              SVNGRKIRMR CRAIDGDVTF CRPKSPVYVG NGVHANLHVA FHRSSSEKIH
 2.5 Å DSSP      EESSSEEEEE EE TTTTB E EEESS  B S SS    EEEE EE  SSS
 2.0 Å DSSP      EETTEEEEEE EEE SSSEEE EEESS EEEB TTB EEEEEE EEES  S
 2.0 Å visual    EETTEEEEEE EEE    EEE EEE   EEE  TT  EEEEEE EEE
                   I'    RB6         RB7      RB7' II      RB8

501              SNEISSDSIG VLGYQKTVDH TKVNSKLSLF FEIKS
 2.5 Å DSSP      S TTTTT TT S  EEEEEEE ETTEEEEEEE    B
 2.0 Å DSSP      TTT   SEEE EEEEEE GGG S EEEEEEEE  EEE
 2.0 Å visual    TT     EEE EEEEEE       EEEEEEEE EEEE
                 I         RB9              RB10
```

**Fig. 8.** Secondary structure assignments in dimeric DT. Upper line, amino acid sequence (Greenfield et al., 1983); second line, DSSP secondary structure assignment for the initial model at 2.5 Å using the nomenclature of Kabsch and Sander (1983): E, $\beta$-strand; H, $\alpha$-helix; G, $3_{10}$-helix; B, isolated $\beta$-bridge; S, bend; T, hydrogen bonded turn; third line, DSSP secondary structure assignment for the refined model at 2.0 Å; fourth line, visual secondary structure assignment for final model at 2.0 Å; fifth line, reverse turns (classified into standard types [Richardson, 1981]) and names of $\beta$-strands and $\alpha$-helices (underlined) using the 3-character code described in the caption to Figure 1. Asterisks indicate residues omitted from the refined model (residues 188–199). Residues at the edges of secondary structures were visually inspected and assigned to the secondary structure only if both $\phi$ and $\psi$ are within the allowed regions of the Ramachandran plot and if the following conditions are met. A residue bordering a helix was included in the helix if it was not in a reverse turn and made at least 1 hydrogen bond to the helix, with donor to acceptor distance $\leq$3.5 Å and the angles at the O and H atoms not less than 110° (Baker & Hubbard, 1984). A residue bordering a $\beta$-strand was included in the strand if it made at least 1 hydrogen bond to a neighboring strand with donor-to-acceptor distance $\leq$3.5 Å and the angles at the O and H atoms not less than 120° (Baker & Hubbard, 1984). Although they are considered $\alpha$-helical by DSSP, residues 352–357 are not included in helix TH9 because residues 352–354 are poorly ordered and residues 355–357 do not form hydrogen bonds with the corresponding $i+4$ residues in TH9. Hydrogen bonded turns were classified into standard types (Richardson, 1981; Wilmot & Thornton, 1990) on the basis of their main-chain torsion angles at the $i+1$ and $i+2$ residues according to the criteria of Lewis et al. (1973). In all turns, the hydrogen bond donor and acceptor atoms of the $i$ and $i+3$ residues are less than 3.3 Å apart with the angles C-O···H and N-H···O >90°.

following the fourth previously assigned helix; a $3_{10}$-helix (CH5′) is assigned preceding the last $\beta$-strand; and the last previously assigned helix (CH6) is deleted. In the T domain: an additional $\alpha$-helix (TH5′) is assigned between the previously assigned fifth and sixth helices. In the R domain: several additional short $\beta$-strands are assigned, RB2′ and RB4′, following the second and fourth previously assigned $\beta$-strands, respectively; and RB3′ and RB7′, which contain residues previously assigned to the fourth and seventh $\beta$-strands, respectively.

*Side-chain conformation*

A combination of side-chain torsion angles ($\chi$ angles) that is highly populated in well-refined structures is termed a rotamer. The fraction of amino acid residues with the $\chi$ angles of a rotamer is termed rotamericity; different amino acids have different rotamericities. Side-chain rotamericity is another basis for comparison of the DT structure with a database of well-refined structures.

Side-chain rotamericity for each amino acid type in DT was compared to the values obtained by Schrauber et al. (1993). The comparison is given in Table 2. Rotamericity values for side chains in DT are generally in the range of the values obtained from the database of well-refined structures. The majority of unusual side-chain conformations observed in dimeric DT can be rationalized by inspection of the model. Several Thr, Phe, and Arg side chains are within 0.2–4.5° of a rotamer boundary; if the criterion of ±20° were slightly relaxed, the rotamericities for these residues would be comparable to the reference values. In addition, Schrauber et al. (1993) note that Asp and Asn have virtually uniform $\chi_2$ distributions, hence low rotamericities. This is consistent with the fact that 17 of 29 Asn residues in DT are excluded from rotamer classes because of unusual $\chi_2$ angles. Asp actually has a 25% higher rotamericity than expected; its high apparent rotamericity may be a consequence of the small number of Asp residues in DT, which occurs at one-half of the expected frequency (Lehninger, 1975).

Frequently residues adopt unusual $\chi$ angles because of tertiary contacts. In DT, 4 tyrosines with unusual rotamers are in tertiary contacts, which may contribute to the low rotamericity of Tyr in Table 2. These tyrosine residues, Tyr 54, Tyr 179, Tyr 375, and Tyr 380 are in the active site, the interface between the C and T domains, the interface between the T and R domains, and the hinge loop, which changes conformation upon dimerization, respectively. Tyr 54 and Tyr 179 adopt unusual $\chi_2$ angles (62° and 295°, respectively), presumably to avoid steric clashes with other atoms. Tyr 380 in the hinge loop is very near the side chain of Lys 385, which may cause its atypical $\chi_2$ angle (305°). Tyr 375 adopts an unusual $\chi_1$ value (277°) in order to position its hydroxyl group in proximity to the main-chain oxygen of residue 481 on the symmetry-related R domain. If $\chi_1$ of Tyr 375 were forced to adopt the $g^-$, $g^+$, or $t$ rotamer, the hydrogen bond would be disrupted.

**Table 2.** *Percentage of side-chain conformations belonging to rotamer classes*

| Amino acid | Number residues[a] | Number rotamers observed/ possible[b] | Rotamericity[c] (%) | Reference rotamericity[d] (%) |
|---|---|---|---|---|
| Valine | 44 | 3/4 | 93.2 | 94.0 |
| Leucine | 35 | 4/9 | 77.1 | 84.1 |
| Isoleucine | 33 | 5/7 | 72.7 | 83.1 |
| Serine | 48 | 3/3 | 83.3 | 85.3 |
| Threonine | 30 | 3/3 | 73.3 | 92.6 |
| Cysteine | 4 | 3/3 | 100.0 | 92.4 |
| Phenylalanine | 18 | 4/5 | 61.1 | 81.9 |
| Tyrosine | 18 | 3/4 | 38.9 | 75.8 |
| Tryptophan | 5 | 2/6 | 80.0 | 78.5 |
| Histidine | 16 | 4/6 | 68.8 | 68.0 |
| Aspartic acid | 28 | 3/3 | 71.4 | 46.1 |
| Asparagine | 29 | 3/6 | 13.8 | 37.5 |
| Glutamic acid | 37 | 6/8 | 54.1 | 69.4 |
| Glutamine | 16 | 3/7 | 62.5 | 73.6 |
| Methionine | 8 | 3/5 | 75.0 | 77.2 |
| Lysine | 39 | 5/7 | 66.7 | 60.7 |
| Arginine | 13 | 2/5 | 23.1 | 65.8 |

[a] Excluded from the table are: Ala (40 residues), Gly (41 residues), Pro (21 residues), residues 188–199.

[b] Side-chain conformations in DT were analyzed and compared to the rotamer library of Schrauber et al. (1993). "Possible" rotamers are unique rotamers determined from a survey of 68 structures of better than 2.0 Å resolution, with $R$ factors below 20.0%. The expanded rotamer library of Schrauber et al. includes a subset of rotamers defined by Ponder and Richards (1987), as well as newly defined rotamers. "Observed" rotamers are unique rotamers found in DT.

[c] Rotamericity is defined as the fraction of residues having side-chain torsion angles within ±20° of the $\chi$ angles of a rotamer.

[d] Reference values for rotamericity as determined by Schrauber et al. (1993).

*Disulfide bridges*

DT contains 4 cysteine residues that form 2 disulfide bridges between Cys 186 in the C domain and Cys 201 in the T domain, and between Cys 461 and Cys 471, both in the R domain (see Kinemage 2). Disulfide bond lengths and angles are given in Table 3.

The Cys 186–Cys 201 disulfide bridge is functionally understood: it is the only remaining covalent bond between the C and T domains after proteolytic cleavage at the arginine-rich proteolysis loop (188–199). The C domain becomes catalytically active when it is freed from the T and R domains by reduction of this disulfide. The $C_\alpha$–$C_\beta$ vectors of the half cystines are antiparallel and the disulfide has a right-handed conformation ($\chi_3 = 78°$).

The Cys 461–Cys 471 disulfide bridge was proposed to be involved in membrane translocation, but its role remains unclear (Papini et al., 1987; Stenmark et al., 1991). It is a rare example of a disulfide bridging 2 antiparallel $\beta$-strands in a protein structure (Fig. 9). Its torsion angles (Table 3) are characteristic of the short right-handed hook conformation, which was observed in crystal structures of peptides that form antiparallel $\beta$-strands with S–S bridges (Karle et al., 1988, 1989). The Cα–Cα distance and S–S bond length are comparable to values found in these peptide structures; however, the CSS angles of 99.6° for Cys 461 and 99.5° for Cys 471 in DT are smaller than the values of 102.9–107.9° determined for these and other linear and cyclic disulfide-containing peptides (Bigoli et al., 1981; Ravi et al., 1983), which may be a consequence of the disulfide occurring in the context of a $\beta$-sheet, rather than in a short peptide with flexible ends.

*Chain flexibility*

The average isotropic $B$ factor for all atoms in the refined dimeric DT model is 29 Å$^2$ (Table 1), which agrees with the value derived from the Wilson plot (see Materials and methods).

**Table 3.** *Disulfide bond parameters*

| Disulfide bridge | $\chi1$[a] (deg) | $\chi2$ (deg) | $\chi3$ (deg) | $\chi2'$ (deg) | $\chi1'$ (deg) | Cα–Cα (Å) | S–S (Å) |
|---|---|---|---|---|---|---|---|
| Cys 186–Cys 201 | 59 | 54 | 78 | 64 | −163 | 5.54 | 2.05 |
| Cys 461–Cys 471 | −55 | −103 | 109 | −82 | −59 | 3.77 | 2.03 |

[a] $\chi1$ and $\chi2$ refer to the first cysteine, and $\chi1'$ and $\chi2'$ to the second cysteine in each pair.

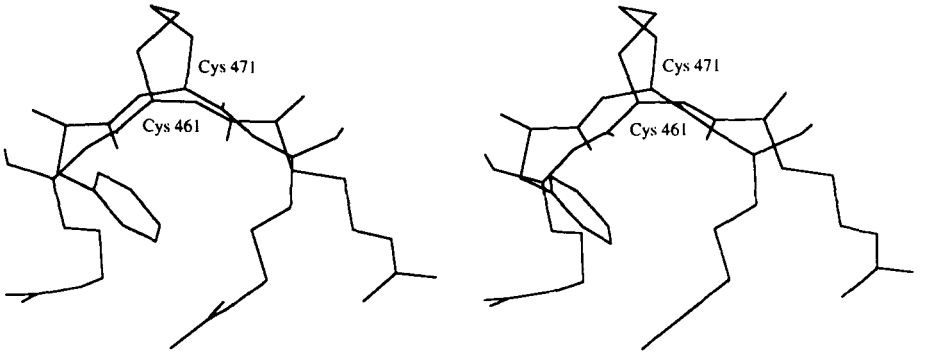


**Fig. 9.** Stereo figure of the disulfide bridge between 2 antiparallel β-strands in the R domain of DT. All non-hydrogen atoms of residues 460–462 and 470–472 are shown. Cysteines 461 and 471, which form the disulfide bridge, are labeled.

Figure 10 shows a plot of the main-chain temperature factors for the 523 residues in the refined model. There are 4 segments with high temperature factors that require further comment. As shown in Figure 10, these are: residues 231–241, 346–354, 464–466, and 518–521.

All residues with high *B* factors in Figure 10 are in surface loops. Residues 231–241 connect the second and third helices in the T domain. Residues 346–354 are in a loop between a pair of helices in the T domain (TH8 and TH9), which may insert into the endosomal membrane at low pH (discussed below). Residues 464–466 are in a loop between the disulfide-linked β-strands RB6 and RB7 (Fig. 1B) in the R domain. Residues 518–521 are in the protruding receptor-binding loop between the last 2 β-strands in the R domain.

*Solvent structure*

The refined model contains 405 water molecules modeled as oxygen atoms. The highest *B* factor for a water molecule is 59 Å$^2$ and the average is 31 Å$^2$. Because the dimeric DT crystals contain 54% solvent based on a protein density of 1.35 g/cm$^3$, the water molecules in the model constitute 14% of the total solvent molecules in the crystal. There are 319 waters in the first hydration shell at a distance of ≤3.5 Å from a polar protein atom; 157 of these can be considered an integral part of the protein, with *B* factors less than or equal to the average *B* factor of 28 Å$^2$ for the main-chain atoms of the polypeptide chain.

Several pentagonal rings of water molecules and polar protein atoms are found in DT, where they are centered over the methylene carbons in long, charged side chains at the surface and in intermolecular contacts. Clusters of water molecules and polar protein atoms in pentagonal rings were first called to attention in the description of the crambin structure (Teeter, 1984), in which they formed caps for apolar atoms. One such cluster in DT has 2 fused pentagons formed by 7 water molecules and Oε of a Gln residue; 1 pentagon is centered on Cδ of Lys 51, and the other on Cγ of Glu 497 in a molecule related by a unit cell translation along *c* (Fig. 11). Two other pentagonal rings are centered on Cγ of Arg 407 and Cγ of Glu 142 in a 2-fold symmetry-related molecule. The methylene carbons make 2 or 3 contacts of less than 4 Å with the water molecules in the rings, except for Cδ of Lys 51, which makes 3 contacts at a distance of 4.2 Å. The average *B* factor for waters in the pentagonal rings is 26 Å$^2$, and the average O–O distance is 2.7 Å. These semiclathrate rings of water around apolar protein groups are consistent with the notion that the hydrophobic interaction is caused by the localization of clathrate-like cages of water about apolar protein groups (Kauzmann, 1959).

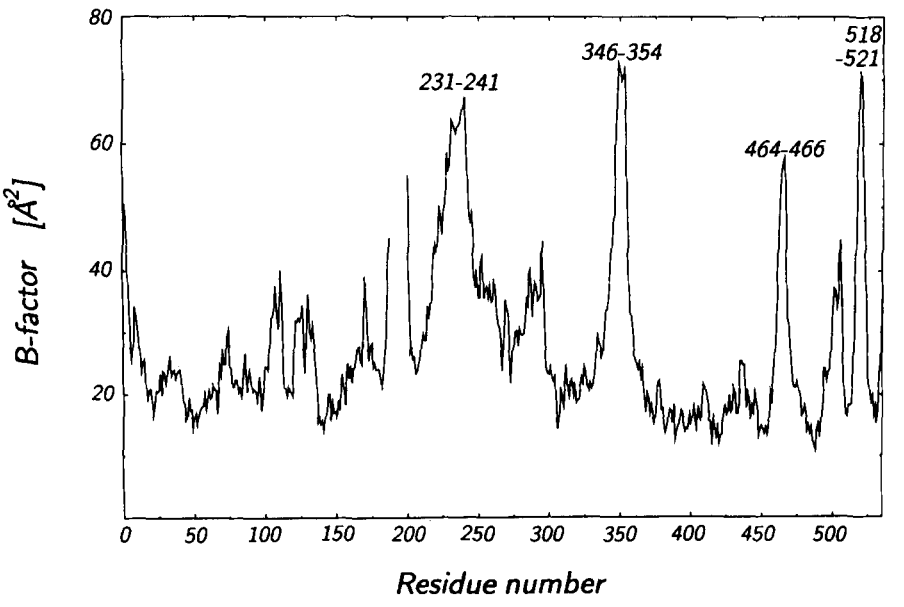


**Fig. 10.** Plot of residue temperature factors. Isotropic main-chain *B* factors, averaged per residue, were plotted against residue number. All 4 labeled segments with high *B* factors are surface loops.

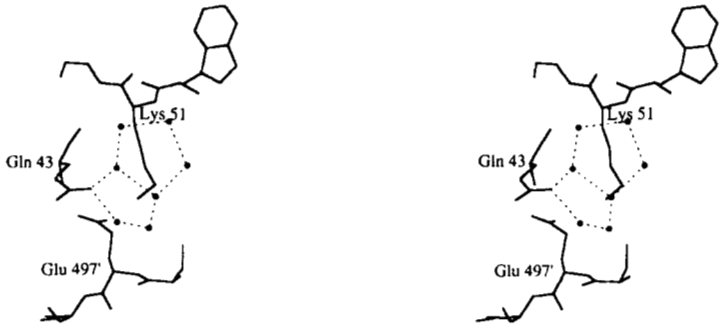


**Fig. 11.** Stereo figure of a semi-clathrate water array in DT. One pentagonal array composed entirely of water molecules is centered over Cδ of Lys 51. Water molecules are represented by their oxygen atoms. A second pentagon, composed of 4 water molecules and a side-chain oxygen atom of Gln 43, is centered over Cγ of Glu 497′ in a c-axis-related molecule (primed residue number indicates symmetry-related residue).

During refinement, an isolated electron density peak of approximately 6 Å by 10 Å was a persistent feature at the dimer interface on the 2-fold axis. Because of the shape of the peak, which was a roughly planar diamond with 4 bulges at the vertices, it could not logically be modeled as a tetrahedral phosphate ion, although phosphate is necessary for dimerization. The peak is not coordinated by the protein as it would be if it were a metal ion. Although the identity of the peak remains obscure, it was modeled as 4 water molecules (2 unique: waters 611 and 735, and their symmetry pairs), which are rather close, the shortest oxygen-to-oxygen distance being 2.4 Å. These waters have *B* factors of 21 and 30 Å$^2$, respectively, and are well-defined in the final $2F_o - F_c$ electron density maps.

*Dinucleotide-binding fold*

Although one of the substrates of DT is NAD, the structure of dimeric DT complexed with the inhibitor ApUp (DT-ApUp) is distinct from the Rossmann NAD-binding fold found in the dehydrogenases (Adams et al., 1970). Figure 7A shows the position of the ApUp molecule in the context of the Cα coordinates of the C domain (see also Kinemage 2). All protein atoms interacting with ApUp, either directly or through a single water molecule, are located in the cleft formed by residues 17–65 in the C domain and residues 446 and 458 in the R domain. The secondary structures formed by these residues comprise the ApUp-binding cleft.

Other bacterial toxins with structurally homologous active sites include heat-labile enterotoxin (Sixma et al., 1991), pertussis toxin (Stein et al., 1994), and exotoxin A (ETA) (Allured et al., 1986). In addition, ETA catalyzes the same reaction as DT, the transfer of ADP-ribose from NAD to diphthamide of EF-2. Superposition of the Cα coordinates of the catalytic domains of DT and ETA yields an RMS difference of 1.56 Å between 91 residues (16–33, 34–38, 46–66, 75–88, 91–96, 133–141, 147–164 of DT; and 435–452, 454–458, 462–482, 493–506, 511–516, 539–547, 552–569 of ETA).

Two previous models for NAD binding to ETA and DT predicted that 1 base of NAD binds in a hydrophobic cavity formed by Tyr 54 and Tyr 65 in DT and the other base stacks against Trp 50 in DT (Brandhuber et al., 1988; Domeninghini et al., 1991). Our model of DT-ApUp, based on an atomic resolution crystal structure, reveals that ApUp is bound roughly perpendicular to the orientation of NAD predicted by these studies. Uracil is stacked against Tyr 65, near Tyr 54; however, adenine is more than 15 Å from Trp 50. It is possible that DT also binds NAD in this orientation because adenosine is present in both ApUp and NAD; to bind NAD in a different orientation would require 2 distinct adenosine-binding sites in DT. Because of the structural similarity of the active sites in DT and ETA, ETA may also bind NAD in this orientation.

*Conformation of bound ApUp*

There are 2 changes in the conformation of the ApUp molecule in the refined model as compared to the initial model (Choe et al., 1992). The adenosine ribose is rotated 180° about χ, the torsion angle around the bond between C1′ in ribose and N9 in adenine, placing C5′ of the ribose and its O5′ oxygen into an electron density peak, and forming a hydrogen bond between O2′ in ribose and His 21. Uracil is also rotated 180° about χ, the torsion angle about the bond between C1′ in ribose and N1 in uracil, placing its O2 atom into an electron density peak and forming hydrogen bonds between N3 and O2 and 2 ordered waters (waters 598 and 577). The entire ApUp molecule is well defined in the electron density with a low average *B* factor of 18 Å$^2$.

The bond lengths and angles of ApUp in DT-ApUp were compared with those in the crystal structure of adenylyl-3′, 5′-uridine (ApU) (Seeman et al., 1976). None of the bonds were found to differ by more than 0.06 Å, and only 2 angles differ by more than 10.0° in the 2 structures. One of these angles is also more than 10° from an equilibrium value used in refinement of DT-ApUp: the angle formed by C5′ and O5′ of uridine and the middle phosphate is 132.5° in DT-ApUp and 118.5° in ApU, as compared to the equilibrium value of 120.5°. The similar angle formed by C3′ and O3′ of uridine and the terminal phosphate (131.7°), for which there is no comparison in ApU, is the only other angle in ApUp more than 10° from an equilibrium value used in refinement. A possible explanation for the deformation of these angles is that it is caused by the salt links between Arg 458 and Lys 24 and the phosphate groups.

As shown in Figure 12A and B, the conformation of ApUp bound to DT resembles an L, in which the adenosine phosphate portion of ApUp is 1 arm of the L and the uridine phosphate portion is the other arm. Table 4 lists selected torsion angles for ApUp in the refined model of DT-ApUp. These angles are within the ranges observed in other crystal structures of nucleotides bound to protein (Moodie & Thornton, 1993). The uridine ribose has a δ angle consistent with the C1′ exo conformer, an energetically acceptable conformer significantly populated in DNA crystal structures. The adenosine ribose in DT-ApUp has the common C2′ endo conformation. As indicated by the values of χ in Table 4, adenosine and uridine have syn and anti conformations, respectively. Unlike NAD bound to the dehydrogenases (Branden & Eklund, 1980), the conformations

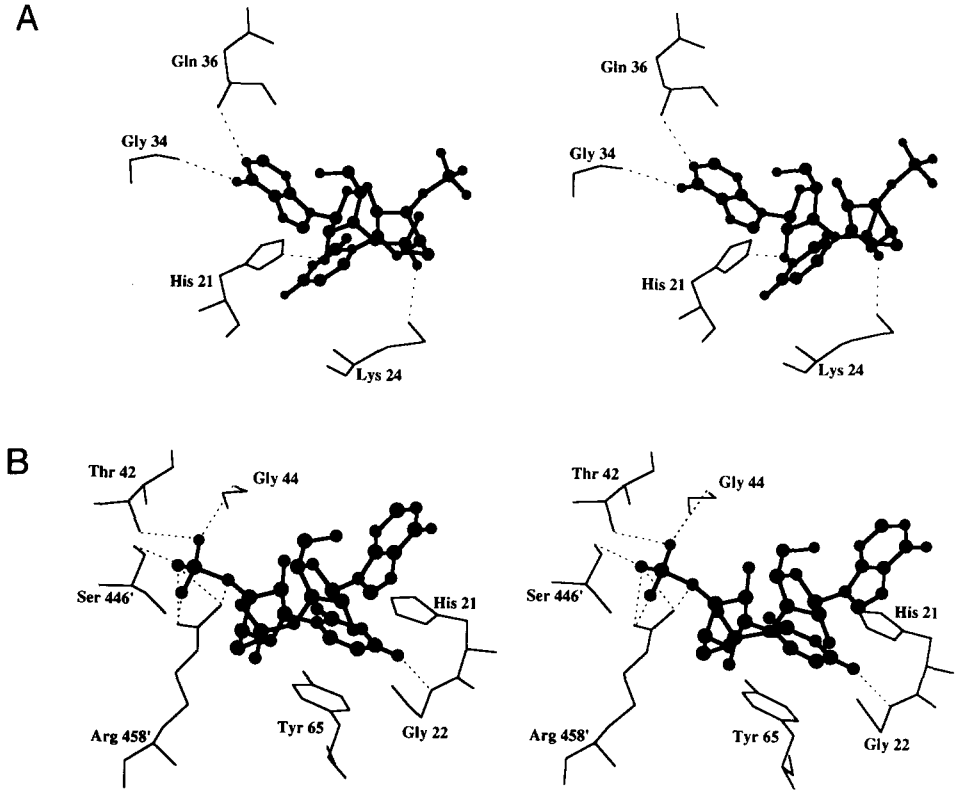


**Fig. 12.** Stereo figures of the active site of DT with the bound nucleotide inhibitor ApUp. **A:** Adenosine phosphate portion of ApUp with the side chains of DT to which it binds. The orientation is approximately the same as Figures 1 and 7. **B:** Uridine phosphate portion of ApUp and the side chains of DT to which it binds. The orientation is rotated approximately 90° about the vertical axis relative to A, as if viewing from the right side of A. The ApUp molecule is shown in ball-and-stick representation with atom sizes proportional to Van der Waals radii. Primed residue numbers indicate residues from the 2-fold symmetry-related R domain. In monomeric DT, the same residues from the R domain are involved in ApUp binding.

about the torsion angles $\gamma$ in ApUp bound to DT are not high-energy conformations.

### *Atomic interactions between DT and ApUp*

The structure of ApUp is related to that of NAD (Fig. 13): each has a 6-membered ring, uracil or nicotinamide, linked to ribose; an adenosine group; and 2 phosphate groups. They differ in the linkage of the 2 nucleosides: in ApUp, 1 phosphate links adenosine to uridine 3′-5′, the other is a 3′ terminal phosphate on the uridine. Obviously, the active site of DT must adjust to accommodate the different structures of NAD or ApUp. However, because the nucleoside portions of NAD and ApUp are similar, we may be able to infer the structure of DT-NAD from that of DT-ApUp. The atomic interactions between ApUp and atoms in DT, direct or mediated by a single water, are summarized schematically in Figure 14 (see Kinemage 2).

**Table 4.** *ApUp torsion angles*

| Torsion angle[a] | Value in adenosine (deg) | Value in uridine (deg) |
|---|---|---|
| $\alpha$ | — | −179 |
| $\beta$ | — | −139 |
| $\gamma$ | 67 | 45 |
| $\delta$ | 147 | 123 |
| $\epsilon$ | −67 | −93 |
| $\zeta$ | −71 | — |
| $\chi$ | 40 | −126 |

[a] Atoms involved in torsion angles: $\alpha$, $O3'_{(n-1)}$-P-O5′-C5′; $\beta$, P-O5′-C5′-C4′; $\gamma$, O5′-C5′-C4′-C3′; $\delta$, C5′-C4′-C3′-O3′; $\epsilon$, C4′-C3′-O3′-P; $\zeta$, C3′-O3′-P-$O5'_{(n+1)}$; $\chi$, O4′-C1′-N1-C2 (pyrimidines), O4′-C1′-N9-C4 (purines).

Looking at the interactions between ApUp and residues in the nucleotide-binding cleft, we find that the location in the model of the His 21 side chain, which forms a hydrogen bond with O2′ in the adenosine ribose (Figs. 12A, 14), is consistent with a His modification study that suggested His 21 is near the adenosine ribose in NAD (Papini et al., 1989). Thus, adenosine in DT-NAD may bind in the same position as in DT-ApUp. His 21 has also been proposed to play a role in catalysis, based on the observation that a titratable group with a p$K$ of 6.2–6.3 (His p$K$ = 6.5) must be protonated for catalysis to occur (Wilson et al., 1990).

Tyr 65, another residue proposed to be involved in NAD binding, makes extensive apolar ring-stacking contacts to uracil in ApUp (Fig. 12B). Tyr 65 was previously identified in the nicotinamide-binding site because both NAD and nicotinamide protect it from being photolabeled (Papini et al., 1991). Thus, it is possible that the nicotinamide ring stacks against Tyr 65 in a similar fashion in DT-NAD.

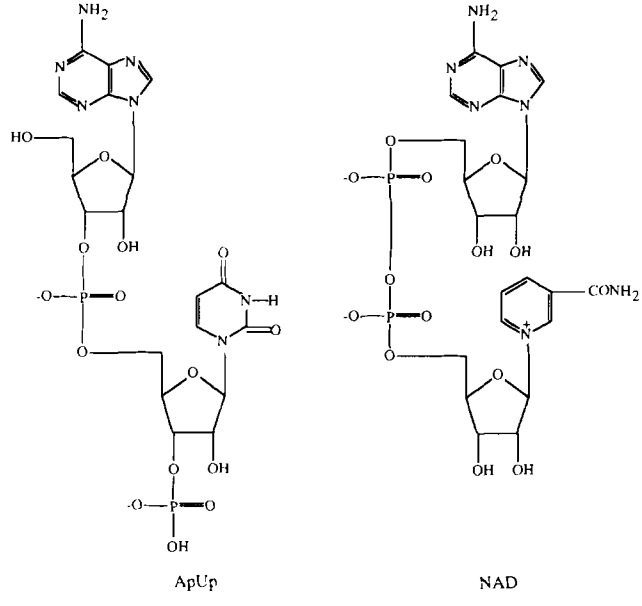


**Fig. 13.** Covalent structures of ApUp and NAD. ApUp is shown on the left and NAD on the right, to illustrate their similarities.

Glu 148 is the only residue in DT that was previously shown to play a role in catalysis; its mutation even to the chemically similar residues Gln or Asp greatly diminishes ADP-ribosylation activity (Wilson et al., 1990). Therefore, Glu 148 is expected to be near the active site. The ApUp atom closest to Glu 148 is O2 in uracil, which is 4.6 Å from a carboxyl oxygen in Glu 148. Photolabeling with nicotinamide-radiolabeled NAD yields nicotinamide linked by its number 6 carbon to the $\gamma$-methylene carbon of Glu 148, suggesting C6 of nicotinamide in DT-NAD is close to Glu 148 (Carroll et al., 1985). If nicotinamide binds in the same position as uracil, C6 will be 5.8 Å from a carboxyl oxygen in Glu 148. Thus, a shared binding site for nicotinamide and uracil is consistent with the photolabeling experiments, although not proven.

A residue that was not previously noted as important in binding NAD or ApUp is Lys 24. In our DT model, Lys 24 makes a salt bridge to the phosphate group between O3′ in adenosine and O5′ in uridine (Figs. 12A, 14). Given the flexibility and length of the Lys side chain, it is conceivable that this salt bridge is present in DT-NAD, even though the phosphate position may be different.

The interaction between ApUp and Ser 446 and Arg 458 in the R domain (Figs. 12B, 14) appears to be specific to DT-ApUp because it involves the 3′ terminal phosphate, which NAD lacks. ApUp interacts with Ser 446 by forming a hydrogen bond, and with Arg 458 by forming a salt bridge and multiple hydrogen bonds between the negatively charged phosphate group and the positively charged guanidinium group. Because the R domains in the DT dimer are swapped, Ser 446 and Arg 458 belong to the 2-fold symmetry-related subunit in the dimer. However, in monomeric DT, the same residues in the R domain are in the active site. The binding of residues in the R domain to the terminal phosphate of ApUp may explain why intact DT binds the substrate NAD (which lacks the terminal phosphate) with micromolar affinity and the inhibitor ApUp with nanomolar affinity, whereas isolated C domain binds NAD 10-fold more tightly than ApUp (Collins & Collier, 1985), and is consistent with the observation that DT binds NADP, with its 2′ phosphate group, more tightly than NAD (Michel & Dirkx, 1974).

The long loop across the active site (residues 34–52) clearly is important for binding nucleotides because several main-chain and side-chain atoms in the loop form hydrogen bonds with ApUp. This active-site loop may be in a different conformation in nucleotide-free DT, in which these interactions are absent. The proposal that the loop undergoes a conformational change upon nucleotide binding is supported by a study that found Lys 39 changes its accessibility to proteases upon NAD or ApUp binding (Zhao & London, 1988). The importance of the loop in nucleotide binding is also suggested by the existence of a mutant DT, known as CRM 197, which has Glu substituted for Gly at position 52 and is unable to bind NAD (Lory et al., 1980). In our structure, Gly 52 is closely packed against a neighboring $\beta$-strand; if the structure is changed by having a side chain at position 52, it may render DT unable to bind NAD or ApUp by repositioning the hydrogen bond donors and acceptors in the active-site loop.

A conformational change in the active-site loop (residues 34–52) upon nucleotide binding could explain the previous observation of changes in the tryptophan band of the near UV CD spectrum in DT-ApUp (Collins & Collier, 1985). In our model, Trp 153 is about 4 Å from adenine, lying in an apolar pocket formed by Lys 37, Gln 43, and Gly 52 in the active-site loop. In nucleotide-free DT, this pocket may not form, altering the environment of Trp 153. Trp 153 is apparently not required for the active-site loop to be in the nucleotide-binding conformation because DT with chemically modified Trp 153 binds NAD, although enzymatic activity is lost (Michel & Dirkx, 1977). The remaining Trp in the C domain, Trp 50 in the active-site loop, is about 10 Å from ApUp. It is possible that one of the tryptophans in DT plays a more intimate role in NAD binding that is not evident in the DT-ApUp complex studied here.

### *Transmembrane domain*

The T domain consists of 10 $\alpha$-helices (Fig. 8). The first 4 helices in the T domain (residues 206–268) lie on the surface of the domain, are hydrophilic, and contain 14 of the 19 basic residues in the entire T domain. In contrast, the last 3 helices in the domain (residues 310–376) are buried in the native structure, are hydrophobic, and contain only 3 positive charges. The middle helices (residues 275–305) are intermediate, in that they are located on the surface of the domain but are hydrophobic and contain only 1 positive charge.

Based on the refined dimeric DT structure and previous hydrophobicity analysis, we suggest that there are 4 potential transmembrane helices in the T domain. Segments 269–289, 301–321, and 338–358 were previously identified from the sequence as potential transmembrane helices based on their mean hydrophobicities using a 21-residue window (Eisenberg et al., 1984). On the basis of the crystal structure, we now suggest that segment 338–358 is not a single transmembrane helix, but includes the most hydrophobic portions of 2 transmembrane helices (residues 328–348 and residues 351–371), which essentially form the last 2 helices in the native DT structure, TH8 and TH9 (Fig. 1B). Segments 328–348 and 351–371 have mean hydrophobicities of

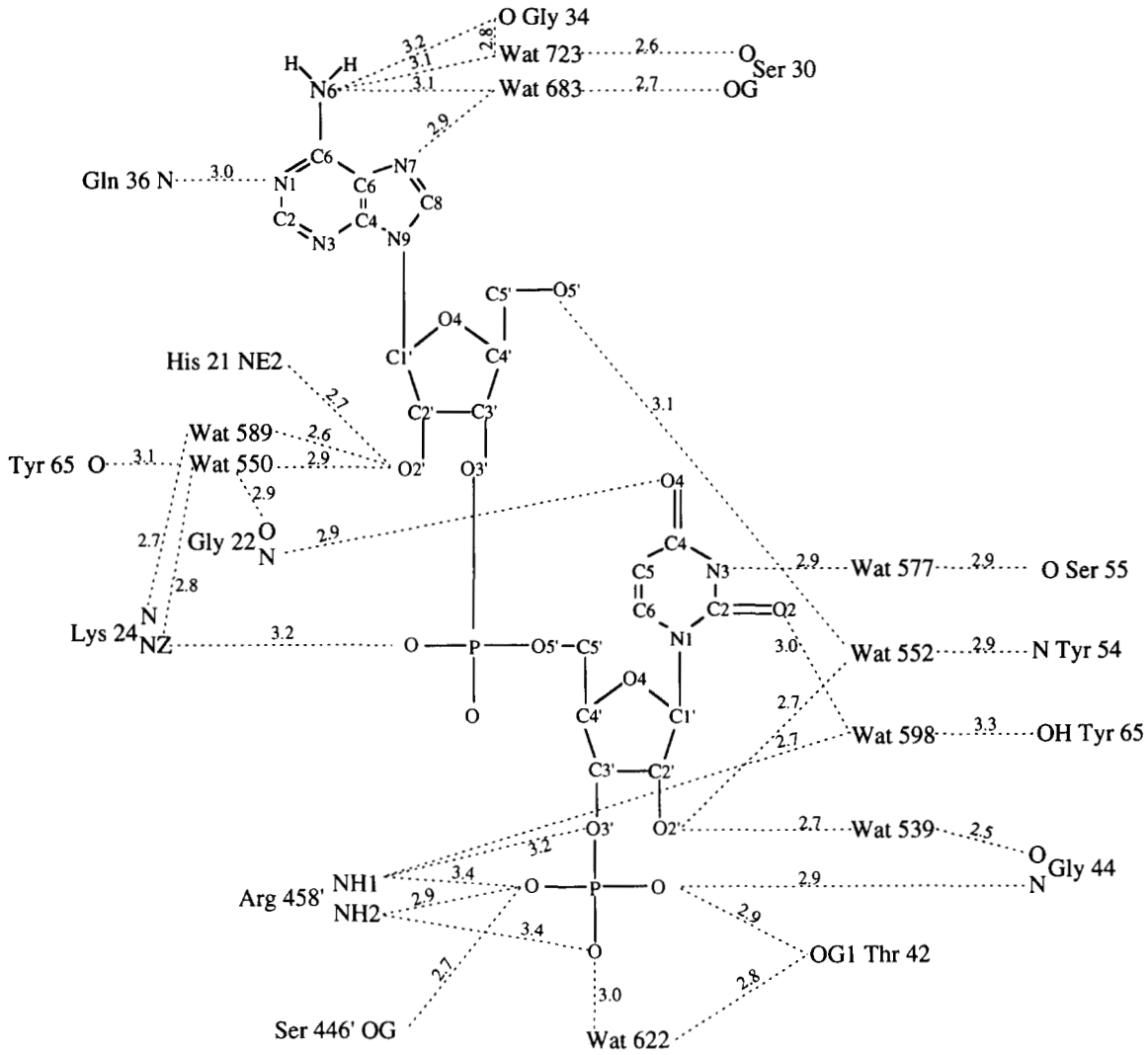


**Fig. 14.** Schematic diagram of the atomic interactions between DT and ApUp in the active site. Hydrogen bonds are shown as broken lines labeled with the donor-to-acceptor distances in Å. Primed residue numbers indicate residues from the 2-fold symmetry-related R domain. In monomeric DT, the same residues from the C and R domains of 1 monomer are involved in ApUp binding.

0.57 and 0.52, well above the threshold value of 0.42 used by Eisenberg et al. (1984) to identify potential transmembrane segments. Thus, combining information from sequence analysis and the crystal structure, we infer that the T domain has 4 potential transmembrane helices (residues 269–289, 301–321, 328–348, and 351–371). These helices do not correspond exactly to helices in the T domain (Fig. 8), but are essentially helices TH5, 6–7, 8, and most of 9 (Fig. 1B); the flanking loops may become helical upon membrane insertion.

DT inserts into membranes when the pH drops (London, 1992), suggesting that low pH increases the hydrophobicity of the T domain. One effect of low pH is to diminish the polarity of protein segments with acidic side chains. Such segments are the 2 loops between sequential pairs of potential transmembrane helices, which hold 6 Asp and Glu residues (Asp 290, Glu 292, Asp 295, Glu 298, Glu 349, and Asp 352), previously referred to as "dagger-tips"; at low pH, protonation of these residues might allow the helices to enter the membrane (Choe et al., 1992). The dimeric DT model also reveals that residues in the 4 potential transmembrane helices are relatively buried in the native structure, with the exception of residues 269–280, 302, 320, and 353. Unfolding of the T domain would expose the buried hydrophobic segments to solvent, an arrangement that would be energetically unfavorable and could cause them to spontaneously insert in the membrane.

The refined dimeric DT model suggests 3 ways that protonation of amino acid residues at low pH could alter the structure and expose apolar segments. First, unfolding of the T domain may be facilitated by 2 nearby histidines, which will repel each other when protonated at around pH 6.5. His 223 and His 257, in the loops following the first and third helices in T, respectively, are only 3.5 Å apart. The mutual repulsion of these His residues may disrupt the packing of these helices, a structural disturbance that might propagate to the rest of the T domain. Because protonation of His residues occurs at a relatively high pH, this may be an initial event in the low pH pathway.

Second, protonating Asp and Glu residues will break salt bridges and may contribute to the unfolding of the T domain. Four of 5 potential intradomain salt bridges in the T domain (Table 5) contain a basic residue from the first helix. One salt bridge links the first helix to the third potential transmembrane helix at the opposite end from the dagger-tip (Lys 212 or Lys 216 to Glu 327). Another salt bridge links the first helix to the fourth potential transmembrane helix (Arg 210 to Glu 362). Breaking these salt bridges could pry the hydrophilic first helix away from the transmembrane helices, exposing them to aqueous solvent and causing them to spontaneously insert into the membrane.

Third, there are at least 3 salt bridges between residues in the first helix in the T domain and in or near the last helix in the C domain. As shown in Table 5 and Figure 15, there is a network of interdomain salt bridges between Asp 207, Arg 210, Asp 211, Lys 172, Arg 173, and Asp 176. An additional salt bridge links Arg 377 at the carboxy-terminus of the T domain and Asp 61 in the C domain. Weakening of these salt bridges at low pH could cause the T and C domains to separate, which may be a prerequisite for membrane insertion.

**Table 5.** *Salt bridges in the transmembrane domain*

| Atom 1[a] | Atom 2 | Distance (Å) |
|---|---|---|
| Intradomain | | |
| Arg 210 NH1 | Glu 362 OE2 | 3.0 |
| Lys 212 NZ | Glu 327 OE1 | 3.0 |
| Lys 216 NZ | Glu 327 OE2 | 2.8 |
| Lys 216 NZ | Glu 259 OE1 | 3.6 |
| Glu 292 OE1 | Lys 299 NZ | 3.6 |
| Interdomain (to C domain) | | |
| Asp 207 OD2 | Lys 172 NZ | 3.7 |
| Asp 207 OD1 | Arg 173 NH1 | 2.6 |
| Arg 210 NH2 | Asp 176 OD1 | 2.8 |
| Asp 211 OD1 | Arg 173 NH2 | 2.5 |
| Arg 377 NH2 | Asp 61 OD1 | 3.0 |

[a] Only salt bridges in which both side chains are well defined in the electron density are included. Where more than 1 interaction between 2 residues is possible, the interaction with the closest distance is given.

*Receptor-binding domain*

The receptor-binding domain of DT can be described as a flattened barrel of 10 antiparallel β-strands with a jellyroll fold (Richardson, 1981). During refinement against the 2.0-Å data, the R domain was significantly changed in 2 areas relative to the initial 2.5-Å model.

The loop connecting the first and second β-strands in R was rebuilt. As we have described previously (Bennett et al., 1994), during refinement of the model at 2.0-Å resolution, it became evident that a 3-residue loop (residues 391–393) had to be changed. In the refined 2.0-Å model, the 391–393 loop covalently links the first β-strand in the R domain to a strand that was previously assigned as the second β-strand in the R domain of the other subunit in the dimer. The first β-strand is now a member of the β-sheet containing the second β-strand (Fig. 1), whereas in the initial model it was not a member of either sheet. This arrangement of β-strands in the R domain of refined dimeric DT is identical to that of monomeric DT (Bennett & Eisenberg, 1994).

The second to last β-strand in the R domain was also changed, by shifting the sequence registration 6 residues toward the carboxy-terminus (discussed above). Residues in this region are involved in binding the DT receptor, and the 6-residue shift in sequence registration forms a β-hairpin loop (residues 514–525) between the last 2 β-strands that protrudes from the surface of the R domain, consistent with a role in molecular recognition, as we discuss in our companion paper on monomeric DT.

*Intermolecular contacts*

Table 6 lists polar interactions at the intermolecular contacts between 1 DT open monomer and its neighbors. The DT dimer dyad is located on the crystallographic 2-fold axis and, because the R domains are swapped, there are many contacts between the C and T domains of one molecule and the R domain of the molecule related by crystallographic 2-fold symmetry (R′). Excluding interactions with ApUp, there are 3 salt bridges, 9 charged and 8 uncharged hydrogen bonds between molecules

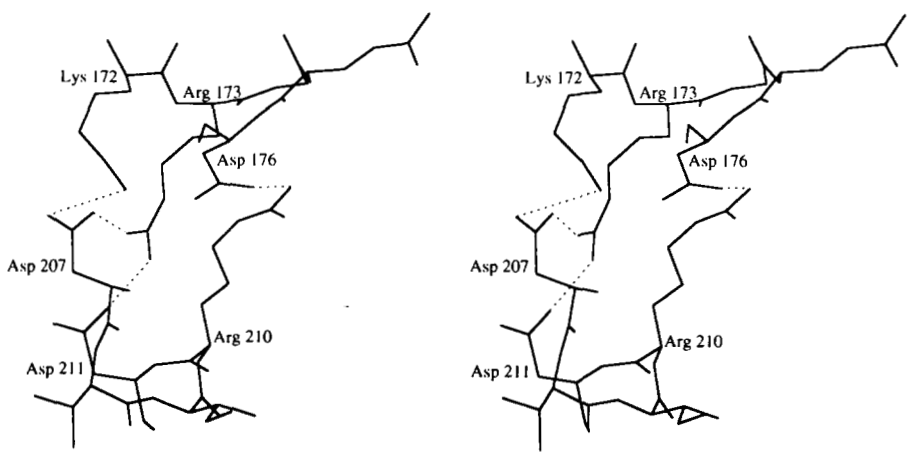


**Fig. 15.** Interdomain salt bridges between the T and C domains. Non-hydrogen atoms of residues 172–176 and 207–211 are shown. Salt bridges (Table 5) are shown as broken lines.

**Table 6.** *Hydrogen bonds and salt bridges formed by an open DT monomer with neighboring molecules in the crystal*

| Atom 1 | | Atom 2[a] | | Symmetry relation[b] | Distance[c] (Å) | |
|---|---|---|---|---|---|---|
| Lys 37 | NZ | Glu 497 | OE2 | I | 3.6 | |
| Lys 39 | O | Lys 440 | NZ | I | 3.5 | |
| Gly 41 | N | Glu 503 | OE2 | I | 3.1 | |
| Gln 43 | NE2 | Glu 497 | OE2 | I | 2.9 | |
| Asn 45 | OD1 | Asn 444 | ND2 | II | 3.1 | C-R′ |
| Tyr 46 | N | Ser 446 | OG | II | 2.9 | |
| Asp 47 | OD2 | His 492 | NE2 | II | 3.7 | |
| Trp 50 | NE1 | Glu 413 | OE2 | II | 3.1 | |
| Asp 97 | OD2 | Lys 447 | NZ | II | 3.2 | |
| Asp 97 | O | Lys 447 | NZ | II | 2.8 | |
| Asn 98 | ND2 | Glu 413 | O | II | 3.0 | |
| Ala 99 | N | Glu 413 | OE2 | II | 3.0 | |
| Phe 140 | O | Lys 456 | NZ | II | 2.8 | |
| Glu 142 | O | Ser 451 | OG | II | 2.8 | |
| Glu 148 | OE2 | Lys 456 | NZ | II | 3.9 | |
| Tyr 179 | OH | Arg 455 | NH1 | II | 3.3 | |
| Gln 287 | O | Lys 522 | NZ | II | 3.5 | T-R′ |
| Ile 306 | O | Gln 515 | N | II | 2.7 | |
| Asp 318 | N | Asn 424 | OD1 | II | 2.7 | |
| Ser 374 | O | Asn 453 | ND2 | II | 2.7 | |
| Tyr 375 | OH | Asn 481 | O | II | 2.7 | |
| Arg 377 | NH1 | Asn 453 | O | II | 3.2 | |
| Tyr 380 | N | His 484 | NE2 | II | 3.0 | Hinge loop-R′ |
| Thr 386 | O | Lys 419 | NZ | II | 2.8 | |
| Ser 446 | OG | ApUp | O3 | II | 2.7 | ApUp-R′ |
| Arg 458 | NH2 | ApUp | O3′ | II | 3.2 | |
| Arg 458 | NH2 | ApUp | O3 | II | 3.4 | |
| Arg 458 | NH1 | ApUp | O1 | II | 3.5 | |
| Arg 458 | NH1 | ApUp | O3 | II | 2.9 | |
| Lys 10 | NZ | Glu 262 | OE2 | III | 3.4 | |
| Asp 57 | OD2 | Lys 264 | NZ | III | 3.1 | |

[a] Atom 2 is in a symmetry-related molecule.

[b] Symmetry operators for the space group C2 relating atom 2 to the corresponding atom in the reference molecule are: I, unit translation along *c*; II, 2-fold axis along *b* at the origin; III, 2-fold axis along *b* at the origin plus *c* centering.

[c] Distance between hydrogen bond donor and acceptor atoms. Table includes all salt bridges ≤4.0 Å and hydrogen bonds ≤3.5 Å.

related by 2-fold symmetry (II). Additional intermolecular contacts in the crystal are between molecules related by unit translation along the *c* axis (I) and by 2-fold symmetry plus *c* centering (III) (Table 6). The contacts between molecules related by I and III include many fewer polar interactions than between molecules related by II.

The solvent-accessible surface area of each pair of molecules forming intermolecular contacts was calculated (Richmond & Richards, 1978). The area buried between molecules related by II is extensive: 4,200 Å$^2$ per molecule, as compared with only 190 Å$^2$ and 570 Å$^2$ between molecules related by I and III, respectively. Our assumption that the dimer in the DT crystals is comprised of 2 molecules related by crystallographic 2-fold symmetry is supported by the extensive area buried and multiple polar interactions between them.

The solvent-accessible surface areas of pairwise combinations of domains forming the 2-fold symmetry contacts were calculated (Richmond & Richards, 1978), subtracted from the sum of solvent-accessible surface areas of the individual domains, and evaluated with atomic solvation parameters (Eisenberg & McLachlan, 1986). The results are shown in Table 7, along with the results for the C-T interface and the R-R′ interface, for comparison. The interface between C and R′ buries 1,860 Å$^2$ and as shown in Table 6, contains 3 salt bridges, and 5 charged hydrogen bonds. As might be expected from the large number of polar interactions in the C-R′ interface, $\Delta G^0_{solvation}$, the free energy change of solvation is only −1 kcal/mol. The interface between T and R′ buries 1,910 Å$^2$ and contains no salt bridges and only 2 charged hydrogen bonds. $\Delta G^0_{solvation}$ upon association of T and R′ is −12 kcal/mol, consistent with the observation that the interface contains apolar segments in both T and R (discussed below).

### *R-R′ interface*

Because DT undergoes domain swapping during dimerization, the dimer interface has 2 classes of interactions. One class includes interdomain interactions that also occur in monomeric DT (C-R′, T-R′). This has been called the "primary interdomain interface" (Bennett et al., 1994). The other class includes interactions found only in dimeric DT (R-R′). This has been called the "secondary interdomain interface" (Bennett et al., 1994).

The secondary interdomain interface (R-R′) is essentially formed by the first 3 β-strands in the R domain and the symmetry-related residues in R′ (Fig. 16). This R-R′ interface consists of 3 vertical layers with alternating apolar/polar character (Fig. 16). The interface buries 440 Å$^2$ per subunit in the dimer (Table 7). If Thr 386, at the end of the hinge loop (Bennett & Eisenberg, 1994), is included in the interface, an additional 35 Å$^2$ are buried and 1 hydrogen bond is formed (Table 6).

All of the polar interactions in the interface are mediated by 1 or more water molecules; 11 of 15 waters in the interface may be considered an integral part of the protein interface, with *B* factors below the average main-chain value of 28 Å$^2$. All R-R′ contacts mediated by a single water are shown in Table 8 and Figure 16. Table 8 and Figure 16 also include polar inter-

**Table 7.** *Solvent-accessible surface area buried between DT domains*

| Domain interaction[a] | Area buried[b] (Å$^2$) | $\Delta G^0_{solvation}$[c] (kcal/mol of monomer) |
|---|---|---|
| C-R′ | 1,860 | −1 |
| T-R′ | 1,910 | −12 |
| R-R′ | 440 | −3 |
| C-T | 1,530 | −1 |

[a] C = catalytic domain, residues 1-187; T = transmembrane domain, residues 200-378; R = receptor-binding domain, residues 387-535, R′ = R domain related by 2-fold crystallographic symmetry.

[b] Area buried per subunit in the dimer as determined by the program ACCESS (Richmond & Richards, 1978).

[c] $\Delta G^0_{solvation}$ is the free energy change of solvation of the surfaces in the domain interface as estimated from atomic solvation parameters (Eisenberg & McLachlan, 1986).

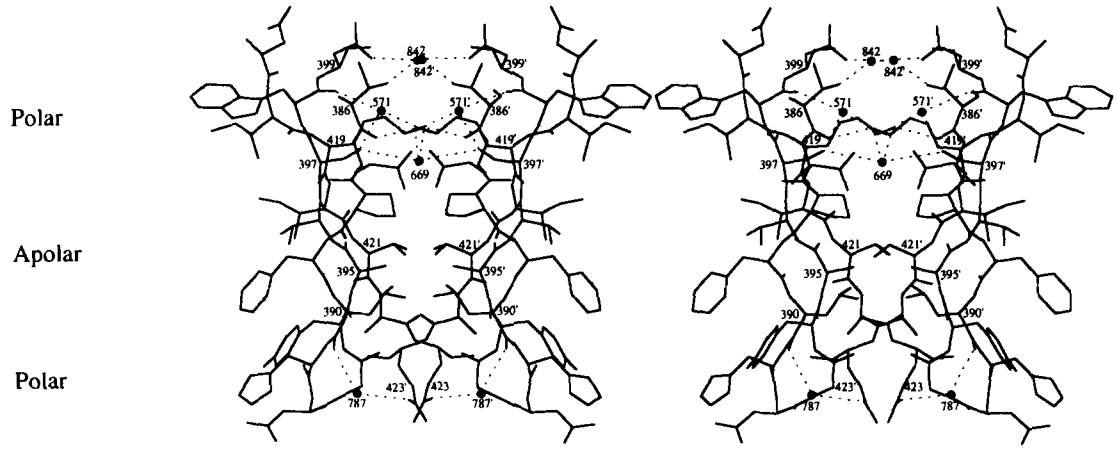


**Fig. 16.** Stereo figure of the dimer interface. Non-hydrogen atoms of residues 386–399 and 417–423 are shown. Water molecules that form hydrogen bonds with residues from both subunits in the dimer are represented by their oxygen atoms (waters 571, 669, 787, and 842). Water molecules hydrogen bonded to other water molecules or to just 1 subunit in the dimer are omitted for clarity. Hydrogen bonds are shown as broken lines. The 2-fold axis is vertical in the plane of the page.

actions between Thr 386 in the hinge loop (Bennett & Eisenberg, 1994) and R′. The first layer of the R-R′ interface, at the bottom in Figure 16, is polar. The unidentified diamond-shaped electron density modeled as 2 waters is near the 2-fold axis (not shown) with protein ligands from the main-chain carbonyl of residue 394 in the second β-strand and the side chain of Glu 423 flanking the third β-strand in R′. A total of 6 more waters (3 unique, 3 symmetry-related) form a cluster with the ring. One of these, water 787, forms a hydrogen bond with residue 390 and residue 423 in R′. The middle layer of the R-R′ interface is apolar, with contacts formed by apolar atoms in side chains from each of the 3 β-strands at the interface (Leu 390, Ala 395, and Thr 421), which interact with the symmetry-related residues in R′. The top layer is polar with contacts formed by residues from 2 β-strands (Ser 397, Trp 398, Asn 399, and Lys 419) and Thr 386 in the hinge loop, which interact through 3 water molecules (waters 571, 669, and 842) with the symmetry-related residues in the other molecule. One of the waters in the top layer sits directly on the 2-fold axis (water 669).

**Table 8.** *Interactions at the DT dimer interface (secondary interdomain interface)*

| Atom 1 | Atom 2 | Distance[a] (Å) | Atom 3[b] | Distance (Å) |
|---|---|---|---|---|
| Top polar layer | | | | |
| Polar contacts (protein) | | | | |
| Thr 386 O | Lys 419′ NZ[c] | 2.8 | | |
| Polar contacts (water) | | | | |
| Thr 386 OG | Wat 842 | 2.4 | Asn 399′ OD1 | 2.9 |
| Ser 397 OG | Wat 669 | 3.6 | Ser 397′ OG | 3.6 |
| Lys 419 NZ | Wat 571 | 3.3 | Trp 398′ O | 3.0 |
| | | | Ser 397′ OG | 2.7 |
| Lys 419 NZ | Wat 669 | 3.4 | Lys 419′ NZ | 3.4 |
| Lys 419 NZ | Wat 669 | 3.4 | Ser 397′ OG | 3.6 |
| Middle apolar layer | | | | |
| Apolar contacts | | | | |
| Leu 390 CD1 | Ala 395′ CB | 3.8 | | |
| Leu 390 CD1 | Thr 421′ CG2 | 4.0 | | |
| Leu 390 CD2 | Ala 395′ CB | 3.8 | | |
| Bottom polar layer | | | | |
| Polar contacts (water) | | | | |
| Leu 390 O | Wat 787 | 3.1 | Glu 423′ OE1 | 3.0 |

[a] Distance between hydrogen bond donor and acceptor atoms (polar interactions) or carbon atoms (apolar interactions).

[b] Atom 1 and Atom 3 from the neighboring subunit in the dimer each form a hydrogen bond with Atom 2 (the oxygen atom of a water molecule).

[c] Primed residue numbers indicate residues in the molecule related by crystallographic 2-fold symmetry.

## Discussion

### *Energetics of dimerization*

Although it is long lived and stable to high salt, guanidinium hydrochloride, and urea, the DT dimer is thermodynamically unstable: dimeric DT dissociates to monomers at a rate of 5–10% per several weeks (Carroll et al., 1986a). This suggests that the binding energy contributed by the dimer interface is insufficient to overcome the loss of entropy upon dimerization. The weakness of association in the dimer is consistent with the fact that dimeric DT is nontoxic (Carroll et al., 1986a): there was no selective pressure for evolving a stable DT dimer.

The dimeric and monomeric DT models reveal structural features that may cause the dimer to be unstable. Because the primary interdomain interfaces (C-R′ and T-R′) are identical in monomeric and dimeric DT, the net area buried upon dimerization is essentially in the secondary interdomain interface (R-R′). Comparison of the solvent-accessible surface areas (Richmond & Richards, 1978) in monomeric and dimeric DT reveals that the area buried in dimeric DT is only 410 Å$^2$ per subunit, consistent with our estimate in Table 7 of the area buried in R-R′. The area buried in the DT dimer is much less than is found in dimers of comparable molecular weight (1,600–4,900 Å$^2$ per subunit) (Janin et al., 1988).

There are no polar interactions in the R–R′ dimer interface and only 2 hydrogen bonds in the hinge loop–R′ interface, fewer than in the hinge loop–R interface in monomeric DT. Thus, the free energy of dimerization may be estimated as the sum of 2 terms: $-T\Delta S^0 + \Delta G^0_{\text{solvation}}$, in which $\Delta S^0$ is the entropy change of rotation and translation and $\Delta G^0_{\text{solvation}}$ is the change in free energy of solvation. We estimate $\Delta G^0_{\text{solvation}}$ from atomic solvation parameters (Eisenberg & McLachlan, 1986) as only −1 kcal/mol of monomer. Because the primary interdomain interfaces (C–R′ and T–R′) are identical in monomeric and dimeric DT, $\Delta G^0_{\text{solvation}}$ reflects the formation of the secondary interdomain dimer interface (R–R′) and changes in the hinge loop (the hinge loop appears to be in a high-energy conformation relative to monomeric DT, as suggested by its low profile scores, discussed above). Following Erickson (1989), we estimate $-T\Delta S^0$ to be 10 kcal/mol of monomer at 277 K, yielding a free energy of dimerization of 9 kcal/mol of monomer. In short, the small R–R′ dimer interface, produced not by natural selection but by freezing in phosphate buffer, does not provide enough binding energy to compensate the loss of entropy upon dimerization.

In spite of its thermodynamic instability, the DT dimer is metastable: it is long lived and resistant to denaturation. What is the source of the metastability of the dimer? We suggest that the barrier to dimer dissociation is the primary interdomain interface formed in dimeric DT by the R domain of one subunit (R′) and the C and T domains of the other subunit. This interface occurs in the toxic monomeric form of DT and presumably evolved over time to be stable. Consistent with stability, the area buried in the primary interdomain interface in the dimer is large (3,770 Å$^2$ per subunit; sum of C–R′ and T–R′ in Table 7), buries apolar surfaces, and contains 3 salt bridges and 15 hydrogen bonds (Table 6). Because the primary interdomain interface is present in both dimeric and monomeric DT, it cannot contribute to the overall thermodynamic stability of dimeric DT. However, disrupting it presents a large activation barrier to dimer dissociation, endowing the dimer with a long lifetime. From the measured rate of dimer dissociation to monomers, we estimate a value of 27 kcal/mol of dimer for the energy of this activation barrier.

### *Implications of the open monomer for toxin function*

Although the domain-swapped dimer is nontoxic, the open monomer within dimeric DT (Fig. 1B) may be relevant to toxin function. Because the domain-swapped dimeric DT structure is formed from the compact, globular monomeric DT structure, we infer that an intermediate must exist in solution, at least transiently, which has altered tertiary structure and may resemble the open monomer within dimeric DT.

Because low pH triggers both the formation of the open monomer structure and membrane insertion, it is possible that the open monomer has structural features in common with a membrane-insertion intermediate. The open monomer is formed by exposing the toxin to low pH, as inferred from the conditions under which dimerization occurs: freezing in mixed phosphate buffer. Mixed phosphate buffer has the property of decreasing from pH 7 to 3.6 during freezing (van den Berg & Rose, 1959), and it appears that low pH is required for dimerization because DT does not dimerize upon freezing in buffers that lack this property (Carroll et al., 1986b).

The open monomer may facilitate membrane insertion by exposing a phosphate-binding site that is buried in monomeric DT. It has been suggested that a phosphate-binding site (*P* site) in DT may be involved in membrane insertion by binding to phospholipids (Lory et al., 1980). A cluster of 9 positive charges partially buried within 6 Å of the interface between the C and R domains may be considered the *P* site of DT (Lys 447, His 449, Arg 455, Lys 456, Arg 458, Arg 460, Lys 474, His 488, and His 492). Lys 474 was previously identified as part of the *P* site on the basis of affinity radiolabeling with ADP-ribose (Proia et al., 1980). Another residue in the cluster, Arg 458, interacts with the 3′ terminal phosphate of ApUp in our DT model. If the *P* site binds phospholipids during membrane insertion, then a more open tertiary structure, perhaps resembling that of the open monomer, must be formed, because the *P* site is buried in the C–R interface in the closed structure of monomeric DT.

In addition to exposing the *P* site, the open monomer has other structural properties that may facilitate membrane insertion. In monomeric DT, the receptor-binding domain, including the protruding β-hairpin loop 514–525, is compact with the T domain (Bennett & Eisenberg, 1994). Therefore, the low pH-triggered conformational change in the T domain leading to membrane insertion could be sterically hindered by the presence of the R domain and possibly the DT receptor. In contrast, the β-hairpin loop 514–525 is 20 Å from the T domain in the open monomer (Fig. 1B).

The open monomer also exposes apolar surfaces that would otherwise be buried in the interface between the T and R domains. Forming the open monomer from monomeric DT requires about 16 kcal/mol in hydrophobic folding energy, as calculated from atomic solvation parameters (Eisenberg & McLachlan, 1986). The unfavorable free energy is the result of exposing apolar residues in the T and R domains. Apolar segments in the T–R interface that are exposed in the open monomer are: residues 306–311 and 316–319 (second proposed transmembrane helix, discussed above), 367–371 (fourth proposed transmembrane helix, discussed above), 426–430, 476, and 483 (in R). Because exposure of these apolar segments to aqueous solvent is unfavorable, they might be expected to lead the insertion of DT into the membrane. The mechanism of cell intoxication by DT is discussed further in the companion paper.

## Materials and methods

### *Purification and crystallization*

Partially purified uncleaved DT was purchased from Connaught Laboratories (Willowdale, Ontario, Canada), incubated with a 2-fold molar excess of the inhibitor ApUp, and further purified into dimeric and monomeric fractions following the method of Carroll et al. (1986b).

Dimeric form IV crystals were grown by the method of Fujii et al. (1991) in 12% (w/v) polyethylene glycol (PEG) 8000, 0.43 M NaCl, 0.043 M Tris-HCl, pH 7.5, using the hanging drop method at 25 °C. Crystal form IV belongs to space group C2 with unit cell parameters $a = 108.3$ Å, $b = 92.3$ Å, $c = 66.1$ Å, $\beta = 90.4$, and 1 molecule per asymmetric unit.

Dimeric DT crystals were prepared for X-ray data collection at −150 °C by soaking in a glycerol-containing artificial mother liquor. Crystals dissolved when 10–30% (v/v) glycerol was added to the crystal growth solution unless the concentration of PEG

8000 was increased from 12% to 18%. Crystals were transferred to 100 μL of artificial mother liquor (18% PEG 8000, 0.43 M NaCl, 0.043 M Tris-HCl, pH 7.5, 20% [v/v] glycerol) in a sitting drop with a 1-mL reservoir of the same composition and soaked for 24 h. Crystals were mounted directly from this solution in a hair loop using the modified method of Teng (1990). Because of the soaking procedure, the unit cell parameters changed to those of crystal form III (Fujii et al., 1991), which are similar to those of crystal form IV. We also observe this unit cell change when form IV crystals are soaked in higher than 14% PEG 8000 in the absence of glycerol; form III crystals are grown from 14% PEG 8000. Upon freezing, the unit cell edges decreased, which is a commonly observed effect (Hope, 1988; Wilke et al., 1991; Madden et al., 1992). The frozen crystal form is referred to as Form V, with unit cell parameters $a = 105.6$ Å, $b = 91.6$ Å, $c = 65.6$ Å, $\beta = 94.6°$.

### *X-ray data collection*

After mounting, crystals were flash frozen and continuously cooled in a −150 °C nitrogen gas stream using an open flow cryostat (Molecular Structure Corporation). Native data were collected from 2 crystals of dimeric DT using an RAXIS-II imaging plate (Rigaku). Eighty-two frames were collected using exposure times of 1 h and 2.1–3° oscillation ranges. The data were integrated and scaled using the RAXIS-II data processing software (Rigaku). The merged data set consisted of 39,447 unique reflections with $R_{merge}$ of 7.3%. The data were 93.0% complete to 2.0 Å resolution. The successful merging of data collected from 2 frozen crystals shows that the change in unit cell dimensions caused by freezing is reproducible. The Wilson plot of these data (Fig. 17) is linear to the high-resolution limit, indicating that data quality is acceptable at 2.0 Å. At 2.0 Å, more than 30% of the reflections had intensities greater than $3\sigma(I)$.

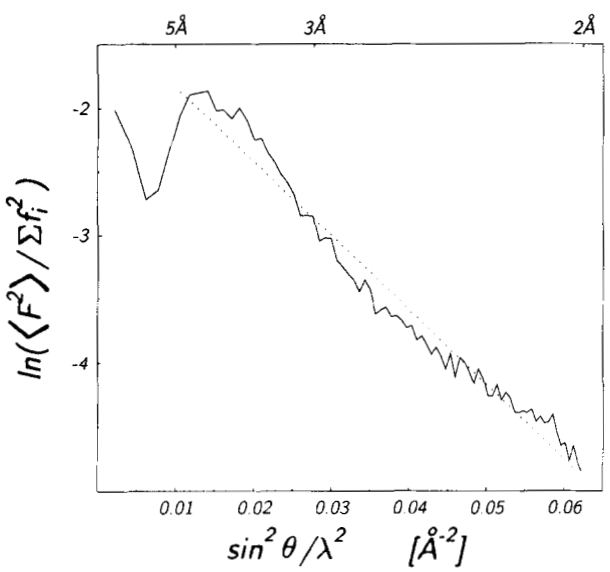


**Fig. 17.** Wilson plot of dimeric DT crystal Form V. The plot (Wilson, 1949) shows the reciprocal space dependence of the ratio of the natural log of the average X-ray intensity ($F^2$) to the sum of the squares of the atomic form factors ($f_i$). The dotted line corresponds to an overall temperature factor of 29 Å$^2$.

The average temperature ($B$) factor estimated from the Wilson plot (dotted line) is 29 Å$^2$.

### *Refinement of the atomic model to X-ray data*

The progress at various stages of refinement is summarized in Table 9. Refinement was initiated using the dimeric DT model determined at 2.5 Å resolution from form III crystals (Choe et al., 1992). All refinement was performed using the program XPLOR (Brünger et al., 1990). The form III coordinates were refined for 20 cycles with the entire molecule modeled as a rigid body, followed by 40 cycles with the 3 domains (C, T, and R) modeled as individual rigid bodies. A typical cycle of refinement included positional, simulated annealing, and, after cycle 7, restrained individual isotropic $B$-factor refinements. After the first cycle of refinement the resolution of the data was extended from 2.5 to 2.4 Å, and after another cycle, the resolution was extended to 2.0 Å. The coordinates were then iteratively rebuilt and refined against $2\sigma$ data between 8 and 2.0 Å resolution. In the last 8 cycles, $1\sigma$ data were used for refinement and in the last 3 cycles, data in the resolution range 10–2.0 Å were used in an effort to resolve additional solvent molecules.

The free $R$ value (Brünger, 1992) is a tool to detect overfitting of the model to the data, and for several refinement cycles, starting at cycle 5, when significant changes were being made

**Table 9.** *Course of refinement of the atomic model of dimeric DT to X-ray data*

| Refinement round | Resolution range (Å) | $R$ factor (%)[a] | Free $R$ factor (%)[b] | Comments |
|---|---|---|---|---|
| 0 | 8–2.5 | 37.4 | | Rigid body refinement; form III model: residues 1–535 plus ApUp; all $2\sigma$ data |
| 1 | 8–2.5 | 30.7 | | Overall $B$ factor |
| 4 | 8–2.0 | 33.0 | | Overall $B$ factor |
| 5 | 8–2.0 | 31.1 | 44.4 | 10% of $2\sigma$ data removed from refinement for free $R$-factor calculation |
| 7 | 8–2.0 | 27.4 | 40.4 | Restrained individual isotropic $B$-factor refinement |
| 17 | 8–2.0 | 20.3 | 27.8 | Residues 188–199 removed from model; 196 water molecules |
| 17.5 | 8–2.0 | 22.6 | | All $2\sigma$ data used in refinement; 196 water molecules |
| 28 | 10–2.0 | 19.5 | | All $1\sigma$ data; final model: residues 1–535 (188–199 missing) plus ApUp and 405 water molecules |

[a] $R = \Sigma(|F_{obs} - F_{calc}|)/\Sigma(F_{obs}) \times 100\%$, where $F_{obs}$ and $F_{calc}$ are the observed and calculated structure factor amplitudes.

[b] Free $R$ factor calculated as above for a randomly chosen set of 10% of the X-ray data ($F_{obs}$), which were excluded from simulated annealing refinement.

to the model, 10% of the data chosen at random were omitted in order to calculate a free $R$ value. These data were never subsequently included in the refinement until after cycle 17. Starting values of $R_{free}$ and $R_{cryst}$ were 44.4% and 31.1%, respectively, decreasing to 27.8% and 20.3% at cycle 17. The decrease in $R_{free}$ and its value of only 7.5% greater than $R_{cryst}$ at cycle 17 shows that extensive rebuilding and the inclusion of 196 well-ordered water molecules (at that stage) were not overfitting the data.

SA omit maps (Hodel et al., 1992) were used to reduce model bias in the electron density maps and to reveal areas of the model needing rebuilding. To scan the structure for questionable areas, 50-residue segments were systematically omitted throughout the entire DT sequence. Only the residues themselves were omitted rather than including an additional 3-5-Å-radius shell around them, which worsened the interpretability of maps. Omitted segments were excluded from both the structure factor calculation and the empirical energy function during refinement and from the structure factor and phase calculations for electron density maps with the coefficients $2F_o - F_c$ and phases calculated from the model. Residues having an atom within 3 Å of the omitted segment were harmonically restrained to their starting positions to prevent them from moving into the omitted region.

After cycle 21, the side-chain orientations of Gln, Asn, and His residues were analyzed to find any carboxamide or imidazole groups needing to be rotated 180°. Because the 2 possible orientations of these side chains are ambiguous in electron density maps, the hydrogen bonding potential and comparison of the $B$ factors of the ambiguous atoms were used to identify 4 His and 5 Gln and Asn side chains, which needed to be changed.

Water molecules were added to the model by building into electron density, which met 3 criteria: first, density was present in the conventional $2F_o - F_c$ electron density map contoured at $1\sigma$; second, density was present in the $F_o - F_c$ map contoured at $3\sigma$; and third, there was at least 1 hydrogen bonding partner (water or protein) within 3.5 Å with reasonable geometry. There are 405 water molecules in the refined model.

## Acknowledgments

We thank Drs. M.S. Weiss and R.J. Collier for discussions and NIH for support. Atomic coordinates and structure factors have been deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, New York 11973 (reference 1DDT).

## References

Adams MJ, Ford GC, Koekoek R, Lentz PJ, McPherson A, Rossmann MG, Smiley IE, Schevitz RW, Wonacott AJ. 1970. Structure of lactate dehydrogenase at 2.8 Å resolution. *Nature 227*:1098–1103.

Allured VS, Collier RJ, Carroll SF, McKay DB. 1986. Structure of exotoxin A of *Pseudomonas aeruginosa* at 3.0 Å resolution. *Proc Natl Acad Sci USA 83*:1320–1324.

Baker EN, Hubbard RE. 1984. Hydrogen bonding in globular proteins. *Prog Biophys Mol Biol 44*:97–179.

Bennett MJ, Choe S, Eisenberg D. 1994. Domain swapping: Entangling alliances between proteins. *Proc Natl Acad Sci USA. 91*:3127–3131.

Bennett MJ, Eisenberg D. 1994. Refined structure of monomeric diphtheria toxin at 2.3 Å resolution. *Protein Sci 3*:1464–1475.

Bigoli F, Lanfranchi M, Leporati E, Nardelli M, Pellinghelli MA. 1981. The structures of DL-2,2′-diamino-4,4′-dithiodibutyric acid and DL-2-amino-2′-ammonio-4,4′-dithiodibutyric acid monohydrogen oxalate. *Acta Crystallogr B 37*:1258–1265.

Blewitt MG, Chung LA, London E. 1985. Effect of pH on the conformation of diphtheria toxin and its implications for membrane penetration. *Biochemistry 24*:5458–5464.

Bowie JU, Luethy R, Eisenberg D. 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science 253*:164–170.

Branden CI, Eklund H. 1980. Structure and mechanism of liver alcohol dehydrogenase, lactate dehydrogenase and glyceraldehyde-3′-phosphate dehydrogenase. *Experientia Suppl 36*:40–84.

Brandhuber BJ, Allured VS, Falbel TG, McKay DB. 1988. Mapping the enzymatic active site of *Pseudomonas aeruginosa* exotoxin A. *Proteins Struct Funct Genet 3*:146–154.

Brünger AT. 1992. Free $R$ value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature 355*:472–475.

Brünger AT, Krukowski A, Erickson JW. 1990. Slow cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallogr A 46*:585–593.

Carroll SF, Barbieri JT, Collier RJ. 1986a. Dimeric form of diphtheria toxin: Purification and characterization. *Biochemistry 25*:2425–2430.

Carroll SF, Barbieri JT, Collier RJ. 1986b. Diphtheria toxin: Purification and properties. *Methods Enzymol 165*:68–76.

Carroll SF, McCloskey JA, Crain PF, Oppenheimer NJ, Marschner TM, Collier RJ. 1985. Photoaffinity labeling of diphtheria toxin fragment A with NAD: Structure of the photoproduct at position 148. *Proc Natl Acad Sci USA 82*:7237–7241.

Choe S, Bennett MJ, Fujii G, Curmi PMG, Kantardjieff KA, Collier RJ, Eisenberg D. 1992. The crystal structure of diphtheria toxin. *Nature 357*:216–222.

Collier RJ. 1975. Diphtheria toxin: Mode of action and structure. *Bacteriol Rev 39*:54–85.

Collins CM, Collier RJ. 1985. Circular dichroism of diphtheria toxin, *Pseudomonas aeruginosa* exotoxin A, and various derivatives. *Biochim Biophys Acta 828*:138–143.

Domeninghini M, Montecucco C, Ripka WC, Rappuoli R. 1991. Computer modelling of the NAD binding site of ADP-ribosylating toxins: Active-site structure and mechanism of NAD binding. *Mol Microbiol 5*:23–31.

Donovan JJ, Simon MI, Draper RK, Montal M. 1981. Diphtheria toxin forms transmembrane channels in planar lipid bilayers. *Proc Natl Acad Sci USA 78*:172–176.

Drazin R, Kandel J, Collier RJ. 1971. Attack by trypsin at a specific site within the intact toxin molecule. *J Biol Chem 246*:1504–1510.

Dumont ME, Richards FM. 1988. The pH-dependent conformational change of diphtheria toxin. *J Biol Chem 263*:2087–2097.

Eisenberg D, McLachlan AD. 1986. Solvation energy in protein folding and binding. *Nature 319*:199–203.

Eisenberg D, Schwarz E, Komaromy M, Wall R. 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J Mol Biol 179*:125–142.

Erickson HP. 1989. Co-operativity in protein-protein association: The structure and stability of the actin filament. *J Mol Biol 206*:465–474.

Freeman VJ. 1951. Studies on the virulence of bacteriophage-infected strains of *Corynebacterium diphtheriae*. *J Bacteriol 61*:675–688.

Fujii G, Choe S, Bennett MJ, Eisenberg D. 1991. Crystallization of diphtheria toxin. *J Mol Biol 222*:861–864.

Greenfield L, Bjorn MJ, Horn G, Fong D, Buck GA, Collier RJ, Kaplan DA. 1983. Nucleotide sequence of the structural gene for diphtheria toxin carried by corynebacteriophage $\beta$. *Proc Natl Acad Sci USA 80*:6853–6857.

Higashiyama S, Lau K, Besner GE, Abraham JA, Klagsbrun M. 1992. Structure of heparin-binding EGF-like growth factor. *J Biol Chem 267*:6205–6212.

Hodel A, Kim SH, Brünger AT. 1992. Model bias in macromolecular crystal structures. *Acta Crystallogr A 48*:851–858.

Hope H. 1988. Cryocrystallography of biological macromolecules: A generally applicable method. *Acta Crystallogr B 44*:22–26.

Janin J, Miller S, Chothia C. 1988. Surface, subunit interfaces and interior of oligomeric proteins. *J Mol Biol 204*:155–164.

Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*:2577–2637.

Kagan BL, Finkelstein A, Colombini M. 1981. Diphtheria toxin fragment forms large pores in phospholipid bilayer membranes. *Proc Natl Acad Sci USA 78*:4950–4954.

Karle IL, Flippen-Anderson JL, Kishore R, Balaram P. 1989. Cystine peptides: Antiparallel $\beta$-sheet conformation of the cyclic biscystine peptide [Boc-Cys-Ala-Cys-$NHCH_3$]$_2$. *Int J Pept Protein Res 34*:37–41.

Karle IL, Kishore R, Raghothama S, Balaram P. 1988. Cyclic cystine pep-

tides. Antiparallel β-sheet conformation for the 20-membered ring in Boc-Cys-Val-Aib-Ala-Leu-Cys-NHMe. *J Am Chem Soc 110*:1958–1963.
Kauzmann W. 1959. Some factors in the interpretation of protein denaturation. *Adv Protein Chem 14*:1–63.
Kraulis P. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24*:946–950.
Lehninger AL. 1975. *Biochemistry*. New York: Worth Publishers Inc.
Lewis PN, Momany FA, Scheraga HA. 1973. Chain reversals in proteins. *Biochim Biophys Acta 303*:211–227.
London E. 1992. Diphtheria toxin: Membrane interaction and membrane translocation. *Biochim Biophys Acta 1113*:25–51.
Lory S, Carroll SF, Collier RJ. 1980. Ligand interactions of diphtheria toxin: Relationships between the NAD site and the *P* site. *J Biol Chem 255*: 12016–12019.
Luzzati V. 1952. Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Crystallogr 5*:802.
Madden DR, Gorga JC, Strominger JL, Wiley DC. 1992. The three-dimensional structure of HLA-B27 at 2.1 Å resolution suggests a general mechanism for tight peptide binding to MHC. *Cell 70*:1035–1048.
Michel A, Dirkx J. 1974. Fluorescence studies of nucleotides binding to diphtheria toxin and its fragment A. *Biochim Biophys Acta 365*:15–27.
Michel A, Dirkx J. 1977. Occurrence of tryptophan in the enzymically active site of diphtheria toxin fragment A. *Biochim Biophys Acta 491*: 286–295.
Moodie SL, Thornton JM. 1993. A study into the effects of protein binding on nucleotide conformation. *Nucleic Acids Res 21*:1369–1380.
Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. 1992. Stereochemical quality of protein structure coordinates. *Proteins Struct Funct Genet 12*:345–364.
Morris RE, Gerstein AS, Bonventre PF, Saelinger CB. 1985. Receptor-mediated entry of diphtheria toxin into monkey kidney (Vero) cells: Electron microscopic evaluation. *Infect Immunol 50*:721–727.
Naglich JG, Metherall JE, Russel DW, Eidels L. 1992. Expression cloning of a diphtheria toxin receptor: Identity with a heparin-binding EGF-like growth factor precursor. *Cell 69*:1051–1061.
Papini E, Santucci A, Schiavo G, Domenighini M, Neri P, Rappuoli R, Montecucco C. 1991. Tyrosine 65 is photolabeled by 8-azidoadenine and 8-azidoadenosine at the NAD binding site of diphtheria toxin. *J Biol Chem 266*:2494–2498.
Papini E, Schiavo G, Sandona D, Rappuoli R, Montecucco C. 1989. Histidine 21 is at the NAD+ binding site of diphtheria toxin. *J Biol Chem 264*:12385–12388.
Papini E, Schiavo G, Tomasi M, Colombatti M, Rappuoli R, Montecucco C. 1987. Lipid interaction of diphtheria toxin and mutants with altered fragment B. *Eur J Biochem 169*:637–644.
Pastan I, Chaudhary V, Fitzgerald DJ. 1992. Recombinant toxins as novel therapeutic agents. *Annu Rev Biochem 61*:331–354.
Ponder JW, Richards FM. 1987. Tertiary templates for proteins. *J Mol Biol 193*:775–791.
Proia RL, Wray SK, Hart DA, Eidels L. 1980. Characterization and affinity labeling of the cationic phosphate-binding (nucleotide binding) peptide located in the receptor-binding region of the B-fragment of diphtheria toxin. *J Biol Chem 255*:12025–12033.
Ravi A. Venkataram Prasad BV, Balaram P. 1983. Cyclic peptide disulfides. Solution and solid state conformation of Boc-Cys-Pro-Aib-Cys-NHMe, a disulfide-bridged peptide helix. *J Am Chem Soc 105*:105–109.
Richardson JS. 1981. The anatomy and taxonomy of protein structure. *Adv Protein Chem 34*:167–338.
Richmond TJ, Richards FM. 1978. Packing of α-helices: Geometrical constraints and contact areas. *J Mol Biol 119*:537–555.
Sandvig K, Olsnes S. 1980. Diphtheria toxin entry into cells is facilitated by low pH. *J Cell Biol 87*:828–832.
Sandvig K, Olsnes S. 1988. Diphtheria toxin-induced channels in Vero cells selective for monovalent cations. *J Biol Chem 263*:12352–12359.
Satow Y, Cohen GH, Padlan EA, Davies DR. 1986. Phosphocholine binding immunoglobulin Fab McPC603. *J Mol Biol 190*:593–604.
Schrauber H, Eisenhaber F, Argos P. 1993. Rotamers: To be or not to be? *J Mol Biol 230*:592–612.
Seeman NC, Rosenberg JM, Suddath FL, Kim JJP, Rich A. 1976. The crystal and molecular structure of sodium adenylyl-3′, 5′-uridine hexahydrate. *J Mol Biol 104*:109–144.
Sixma TK, Pronk SE, Kalk KH, Wartna ES, van Zanten BAM, Witholt B, Hol W. 1991. Crystal structure of a cholera toxin-related heat-labile enterotoxin from *E. coli*. *Nature 351*:371–377.
Stein PE, Boodhoo A, Armstrong GD, Cockle SA, Klein MH, Read R. 1994. The crystal structure of pertussis toxin. *Structure 2*:45–57.
Stenmark H, Olsnes S, Madshus IH. 1991. Elimination of the disulphide bridge in fragment B of diphtheria toxin: Effect on membrane insertion, channel formation, and ATP binding. *Mol Microbiol 5*:595–606.
Teeter MM. 1984. Water structure of a hydrophobic protein at atomic resolution: Pentagon rings of water molecules in crystals of crambin. *Proc Natl Acad Sci USA 81*:6014–6018.
Teng TY. 1990. Mounting of crystals for macromolecular crystallography in a free-standing thin film. *J Appl Crystallogr 23*:387–391.
Uchida T, Gill DM, Pappenheimer AM. 1971. Mutation in the structural gene for diphtheria toxin carried by temperate phage β. *Nature New Biol 233*:8–11.
van den Berg L, Rose D. 1959. Effect of freezing on the pH and composition of sodium and potassium phosphate solutions: The reciprocal system $KH_2PO_4$-$Na_2HPO_4$-$H_2O$. *Arch Biochem Biophys 81*:319–329.
Van Ness BG, Howard JB, Bodley JW. 1980. ADP-ribosylation of elongation factor 2 by diphtheria toxin. *J Biol Chem 255*:10710–10716.
Wilke ME, Higaki JN, Craik CS, Fletterick RJ. 1991. Crystal structure of rat trypsin-S195C at −150 °C. *J Mol Biol 219*:511–523.
Wilmot CM, Thornton JM. 1990. β-Turns and their distortions: A proposed new nomenclature. *Protein Eng 3*:479–493.
Wilson AJC. 1949. The probability distribution of X-ray intensities. *Acta Crystallogr 2*:318–321.
Wilson BA, Reich KA, Weinstein BR, Collier RJ. 1990. Active-site mutations of diphtheria toxin: Effects of replacing glutamic acid-148 with aspartic acid, glutamine, or serine. *Biochemistry 29*:8643–8651.
Zhang KYJ, Eisenberg D. 1994. The 3D profile method using residue preference as a continuous function of its environments. *Protein Sci 3*: 687–695.
Zhao JM, London E. 1988. Localization of the active site of diphtheria toxin. *Biochemistry 27*:3398–3403.