Domain swapping: Entangling alliances between proteins
(diphtheria toxin/evolution of oligomers)

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

Molecular Biology Institute, Department of Chemistry and Biochemistry, and University of California, Los Angeles–U.S. Department of Energy Laboratory of Structural Biology and Molecular Medicine, University of California, Los Angeles, CA 90024-1570

Contribution by David Eisenberg, December 27, 1993

ABSTRACT The comparison of monomeric and dimeric diphtheria toxin (DT) reveals a mode for protein association which we call domain swapping. The structure of dimeric DT has been extensively refined against data to 2.0-Å resolution and a three-residue loop has been corrected as compared with our published 2.5-Å-resolution structure. The monomeric DT structure has also been determined, at 2.3-Å resolution. Monomeric DT is a Y-shaped molecule with three domains: catalytic (C), transmembrane (T), and receptor binding (R). Upon freezing in phosphate buffer, DT forms a long-lived, metastable dimer. The protein chain tracing discloses that upon dimerization an unprecedented conformational rearrangement occurs: the entire R domain from each molecule of the dimer is exchanged for the R domain from the other. This involves breaking the noncovalent interactions between the R domain and the C and T domains, rotating the R domain by 180° with atomic movements up to 65 Å, and re-forming the same noncovalent interactions between the R domain and the C and T domains of the other chain of the dimer. This conformational transition explains the long life and metastability of the DT dimer. Several other intertwined, dimeric protein structures satisfy our definition of domain swapping and suggest that domain swapping may be the molecular mechanism for evolution of these oligomers and possibly of oligomeric proteins in general.

Diphtheria toxin (DT) is a 535-residue protein that causes diphtheria. Its pathogenesis involves three functions: receptor binding, membrane translocation, and catalysis, each of which is associated with one of three folding domains (1). After binding to a receptor on eukaryotic cells (2), DT is endocytosed and triggered by low pH to insert into the membrane and translocate its toxic C domain into the cytosol, where it inactivates a factor essential to protein synthesis (3–5).

We previously reported the multiple isomorphous replacement (MIR) structure of dimeric DT at 2.5-Å resolution (1). Now we have extended the x-ray data of this crystal form to 2.0-Å resolution (Table 1), which allowed us to interpret a previously obscure segment of the electron density.5 We have also crystallized monomeric DT in another space group and determined its structure at 2.3-Å resolution by molecular replacement.

MATERIALS AND METHODS

The dimerization of purified monomeric DT bound to the inhibitor adenyl(3′-5′)-uridine 3′-monophosphate (ApUp) was carried out (6) by freezing in 25 mM mixed phosphate buffer with 150 mM NaCl for 24 hr. Dimer and monomer portions were purified by size-exclusion HPLC. Dimeric form IV crystals in the space group C2 were grown by the method of Fujii et al. (7). Dimer crystals were prepared for low-temperature (−150°C) data collection by soaking in artificial mother liquor plus 20% glycerol for 24 hr and were mounted directly from this solution by the modified method of Teng (8). Due to soaking, the unit cell parameters changed to those of crystal form III (also C2) (7), and upon freezing, the unit cell dimensions decreased, so the frozen crystal form is denoted form V, with unit cell parameters a = 105.6 Å, b = 91.6 Å, c = 65.6 Å, and β = 94.6°. Table 1 summarizes the x-ray data. Molecular replacement using the published 2.5-Å dimer model and refinement were performed with programs in XPLOR (9). The refined dimer model has a crystallographic R factor of 19.5% for 37,727 reflections > 1σ (90% complete between 10 and 2 Å); root-mean-square deviation from ideal bond lengths is 0.018 Å and from bond angles is 2.6°. Details will be published elsewhere.

Monomer crystals in the space group P21212 were grown by seeding in conditions similar to those previously described (10). They were prepared for low-temperature data collection by soaking in artificial mother liquor plus 10% glycerol for 4 hr and mounted as described above. Molecular replacement using the 2.0-Å refined dimer model and refinement were performed with programs in XPLOR (9). The refined monomer model has a crystallographic R factor of 20.8% for 42,855 reflections > 1σ (89% complete between 10 and 2.3 Å). Details will be published elsewhere.

RESULTS

Monomeric DT has essentially the structure we reported earlier (1) for the monomer within dimeric DT. However, the refined dimer structure has a small but significant difference. During refinement of the dimer model against the 2.0-Å data, it became evident that a three-residue loop (residues 391–393) had to be changed (highlighted in red in Fig. 1a). This loop follows the first β-strand in the R domain. It is 10 Å from the crystallographic twofold axis and was poorly defined in the MIR electron density. The new path of loop 391–393 covers residues 391–393.

*Present address: Salk Institute, La Jolla, CA 92037.
†To whom reprint requests should be addressed.
5The atomic coordinates and structure factors have been deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973.

<table>
<thead>
<tr>
<th>Crystal</th>
<th>Temperature</th>
<th>No. of crystals</th>
<th>No. of reflections</th>
<th>% Rsym(I)*</th>
<th>% complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimer</td>
<td>−150°C</td>
<td>2</td>
<td>2.0</td>
<td>39,447</td>
<td>93.0</td>
</tr>
<tr>
<td>Monomer</td>
<td>−150°C</td>
<td>2</td>
<td>2.3</td>
<td>44,070</td>
<td>89.8</td>
</tr>
</tbody>
</table>

*Rsym(I) = 100 × (Σ Σ |I(hkl)| − |I(hkl)|)/Σ Σ |I(hkl)|, where I(hkl) is the ith observation of the intensity of the hkl reflection and |I(hkl)| is the mean intensity of the hkl reflection.

Abbreviations: BS, bovine seminal; DT, diphtheria toxin; GM-CSF, granulocyte/macrophage-colony-stimulating factor; IFN, interferon; IL, interleukin; MIR, multiple isomorphous replacement.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
lently links the T domain to the distant R domain (Fig. 1a); in
the old model (1), this loop covalently linked the T domain to
the near R domain. Specifically, the loop 391–393 connects
the first blue β-strand to the second β-strand (blue) of the distant
R domain, whereas in the original model it connected the first
blue β-strand to the second β-strand (yellow) of the near R
domain. The positions of all main-chain atoms in the structure
(other than the three-residue loop) are essentially the same as
in our original dimer model; however, changing this loop
results in a large difference in the position of the R domain
relative to the other two domains of the same molecule. The
loop 391–393 has low main-chain temperature factors and is
extremely well defined in the 2.0-Å electron density map (Fig.
2a and b). A heavy-atom binding site at His391 provides
independent crystallographic evidence for the path of this loop
(Fig. 2c).

Comparison of monomeric (Fig. 1b) and dimeric (Fig. 1a)
 DT shows their similarity. The structures of the three do-
 mains (C, T, and R) are identical, as are the relative positions
of C, T, and R in the monomer and C, T, and R' in the dimer,
where R' indicates the R domain of the neighboring chain.
In other words, the interactions between the C, T, and R
domains of a single polypeptide chain in monomeric DT are
reproduced exactly in dimeric DT by the C and T domains of
one polypeptide chain and the R domain of the neighboring
chain in the dimer. We refer to this exchange of R domains
in dimeric DT as domain swapping. A general definition of
domain swapping is given in the Discussion.

One monomer within the dimeric DT crystal structure can
be described as an "open" monomer (Fig. 1a), in contrast to
the "closed" monomer observed in the monomeric DT crystal
structure (Fig. 1b). The existence of both an open and a closed
form of the DT monomer necessitates a large conformational
rearrangement. This rearrangement occurs by changing the
main-chain conformations of residues 380–386 in the loop
preceding the first β-strand in the R domain (Fig. 3) (highlighted in red in Fig. 1b), which allows the rotation of the entire
15-kDa R domain by 180°.

**FIG. 1.** Relationship of domain-swapped dimeric DT to monomeric DT. (a) Ribbon drawing of dimeric DT based
on the 2.0-Å structure. One polypeptide chain is blue, the other yellow. Loop 391–393 (between the first and second
β-strands in the R domain), which has changed from our original model, is in red. C (catalytic), T (transmembrane),
and R (receptor binding) domains are indicated. Due to disorder, residues 188–199, which form a surface loop linking
the carboxyl terminus of the C domain to the amino terminus of the T domain, are missing from the model. (b) Ribbon
drawing of monomeric DT based on the 2.3-Å structure. Loop 380–386 (preceding the first β-strand in the R domain),
which differs between monomer and dimer, is in red. Ribbon drawings were made with MOLSCRIPT (11).

**FIG. 2.** Electron density maps showing the crossing of the polypeptide chain to the distant R domain. (a) Loop follows-
ing the first β-strand in the R domain (residues 391–393). Orientation is approximately the same as Fig. 1a. The model-phased 2|Fo| − |Fc| electron density map was calculated with phases computed from the partially refined form V coordinates (before rebuilding residues 391–393) and form V experimental
amplitudes. The partially refined form V coordinates superimposed on the map contoured at 1σ show residues from the first β-strand of the R domain on the left and from the second strand of the distant R domain on the right. Because the loop connecting these two strands had not been
modeled before phase computation, this map is essentially an omit map of this segment (that is, unphased model). (b) Refined
form V coordinates (after rebuilding residues 391–393) superimposed on the same map as a. The model fits the electron density well,
including the side chains of His391 and Asp392 and the lack of a side chain on Gly393. This excellent fit of model to electron density confirms
that the polypeptide chain from the T domain traverses to the distant R domain. (c) Osmium binding site at His391 in the loop following the first
β-strand. The occurrence of the osmium atom 2.5 Å from Nε of His391 confirms the model. In the original model, with its incorrect loop, His391
was 7 Å from the osmium atom. The model-phased |Fo| − |Fc| Fourier difference map is shown contoured at 10σ for the heavy-atom derivative
K₂OsO₄ (1), where |Fo| and phases are from form IV data and the refined form V model placed in the form IV unit cell, respectively. The
excellent fit of the model to the electron density is corroborated by the low average main-chain temperature factor in this loop, 16 Å², compared
with 28 Å² for the whole molecule.
DISCUSSION

Monomeric and Dimeric DT Structures. A large conformational change during DT dimerization is consistent with biochemical observations. Monomeric DT does not spontaneously convert to dimer at neutral pH, even at high concentrations (30 mg/ml), indicating a kinetic barrier to dimerization. However, DT can be induced to dimerize by freezing the protein in mixed phosphate buffers, which are known to decrease in pH from 7.0 to 3.6 during freezing (12). Because freezing causes dimerization only in the presence of mixed phosphate buffer and other buffers which similarly decrease in pH upon freezing, Carroll et al. (6) proposed that the decrease in pH causes dimerization. Based on the comparison of the monomeric and dimeric DT crystal structures we propose that the decrease in pH converts monomeric DT to an open form which then dimerizes by domain swapping at the high concentrations of the eutectic mixture as the pH returns to neutral during thawing (Fig. 4).

The monomeric DT structure suggests how low pH triggers open-monomer formation. The interdomain interface between the R and C domains is unusually charged and polarized. There are nine basic and only three acidic residues on the R-domain interface surface and seven acidic residues on the C-domain surface. Three salt bridges stabilize the interface at neutral pH. At low pH, these salt bridges will be disrupted due to protonation of the acidic residues and there will be further destabilization due to isolated, buried positive charges in the interface, favoring the formation of an open monomer structure.

During dimerization of DT, we also observe higher oligomers by size-exclusion HPLC. These higher oligomers include not only tetramers but also trimers and pentamers, which could not be formed simply by dimers associating. It is conceivably that these higher oligomers are formed by domain swapping, making linear or cyclized aggregates in which each molecule interacts with two neighboring molecules by providing an R domain to one and accepting an R domain from the other. The aggregates of other proteins caused by mildly denaturing conditions could conceivably form by domain swapping.

Domain Swapping. In discussing domain swapping, it is convenient to define two types of protein interfaces: primary and secondary interdomain interfaces. We define the primary interdomain interface to be the contacts between domains in the monomeric form of the protein (Fig. 5, I). We consider an oligomer to be domain-swapped when an interface identical to a primary interdomain interface is formed by different polypep-
tide chains. For example, in a domain-swapped dimer, two primary interdomain interfaces are formed between domains from two polypeptide chains (Fig. 5, II). There is also the possibility of forming an additional interdomain interface in the domain-swapped dimer, which we define as the secondary interdomain interface, which is not present in the monomer (Fig. 5, III). Notice that for a dimer to qualify as domain-swapped, a monomeric form of the protein must also be known.

Domain swapping appears to be a mode for oligomerization of other proteins as well as DT. Examples of domain swapping are found in the RNAses (Table 2). BS RNase can be isolated as a monomer, dimer, or domain-swapped dimer, in which the swapped domain is a 15-residue N-terminal segment (19). The attractive forces between the N-terminal segment and the subunit bodies are conserved in all three forms. In a study of the dimerization of BS RNase, the authors stated that, “the main events of the transformation appear to be (i) the destruction of those bonds and forces, (ii) the displacement of the freed segment, hinging on the connecting loop, and (iii) the integral reconstruction of the same bonds and forces, only on a different subunit body” (22). Thus, although the swapped domain is only 15 residues, the BS RNase dimer fits our definition of domain swapping. All three forms of BS RNase are active, but only the swapped dimer exhibits allosteric properties (22).

Similarly, 30 years ago, Crestfield et al. (18) found that RNase A forms a dimer under certain conditions by exchanging its N-terminal segment in a manner that satisfies the definition of domain swapping. The conditions for forming domain-swapped RNase A dimers are reminiscent of the conditions for dimerizing DT: as with DT, RNase A is believed to form an open monomer in low pH. Specifically, the dissociation of the N-terminal segment of RNase A occurs by treatment with 50% acetic acid; low pH is known to induce the separation of the S peptide from the S protein of RNase S. Then, according to Crestfield et al., “Upon cooling the [RNase A] solution to ~80° and lyophilization, some of the dissociated ends are assumed to become bound so stably that disaggregation does not readily occur at room temperature” (18). Because the active site is formed by residues of the N-terminal segment and the body of the RNase A protein, biochemical experiments established that the “NH2-terminal peptide sections mutually displace one another in their binding to the rest of the protein” (29). Thus, the RNase A dimer also fits our definition of domain swapping.

The Domain-Swapping Hypothesis for Evolution of Protein Oligomers. Domain swapping may also be a piece in the puzzle (30) of how oligomeric proteins evolved. The solvent-accessible surface areas of dimer interfaces range from ~700 Å² to ~5000 Å² per subunit (31). The formation of such an extensive complementary interface might require several simultaneous mutations and thus is an unlikely event. This difficulty disappears if evolution of an oligomer involved domain swapping. Domain swapping provides a pre-evolved, tight protein interface (the primary interdomain interface) for forming a dimer interface. The process of domain swapping creates a dimer having the same primary interdomain interface as the monomers. If dimers and higher oligomers evolved by domain swapping, it would be consistent with Miller’s observations that oligomeric interfaces are similar to monomer interiors structurally and in their atomic hydrophathy (33). The domain-swapped dimer may also confer some advantage to the organism such as allosteric properties, as in BS RNase (22). If so, random mutations which enhance the interaction at a secondary interdomain interface will be favored by natural selection, which in time will create a more stable dimer (see Fig. 5, III, and dashed right-hand minimum in Fig. 4). Thus, the pathway (Fig. 5, I, II, III) represents the hypothetical evolution of a stable dimeric protein that passed through a marginally stable domain-swapped dimer.

Such a pathway could have been the genesis of the intertwined dimeric proteins IFN-γ, IL-5, and βB2.

Table 2. Domain swapping in proteins

<table>
<thead>
<tr>
<th>Domain-swapped dimer</th>
<th>Monomer or monomeric homolog</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein (ref.)</td>
<td>Structure determination</td>
</tr>
<tr>
<td>DT dimer</td>
<td>X-ray</td>
</tr>
<tr>
<td>RNase A dimer (18)</td>
<td>Biochemical</td>
</tr>
<tr>
<td>BS RNase dimer (19)</td>
<td>X-ray</td>
</tr>
<tr>
<td>βB2-crystallin (20)</td>
<td>X-ray</td>
</tr>
<tr>
<td>IFN-γ (16)</td>
<td>X-ray</td>
</tr>
<tr>
<td>IL-5 (17)</td>
<td>X-ray</td>
</tr>
</tbody>
</table>

BS, bovine seminal; IFN, interferon; IL, interleukin; GM-CSF, granulocyte-macrophage-colony-stimulating factor.

*aOther intertwined dimer structures without known monomeric forms or monomeric homologs, but which could conceivably have formed by domain swapping, include: trp repressor (26), urotensin (27), and bacteriophage MS2 coat protein (28).

A monomeric homolog is a different protein which has a protein fold highly similar to that of one half of the dimer.
crystallin (20) (Table 2). Each of these dimers has an interdomain interface similar to the primary interdomain interface found in a related, but not identical, monomeric protein. For example, dimeric IL-5 has an interface similar to a primary interdomain interface of the monomeric protein GM-CSF, which has a four-helix cytokine fold (25). IL-5 differs from GM-CSF in having a short loop between helices C and D, which seems to make it impossible for IL-5 to form a closed monomer with the structure of GM-CSF. This shortened loop may have been the genetic change that created domain-swapped dimers of IL-5 from a monomeric protein resembling GM-CSF. IFN-γ and ββ2-crystallin could also have evolved from monomeric to dimeric form by domain swapping. In other words, the original monomeric structures of these proteins, which may have resembled IFN-β (24) and γB-crystallin (23), respectively, are no longer seen.

In addition to these highly intertwined dimers, other present-day oligomers could have evolved from monomeric proteins by passing through a domain-swapped stage (as in Fig. 5, II) on the way to a stable oligomer (Fig. 5, III). These present-day oligomers would not disclose their prior histories of domain swapping, because in evolving from the stage represented in Fig. 5, II, to that of Fig. 5, III, they would have evolved extensive secondary domain interfaces. Also, their original monomeric forms would no longer be stable with respect to their oligomeric forms. In short, domain swapping could have provided a mechanism for evolution of oligomers, in which each subunit adheres strongly to others through extensive "pre-evolved" interfaces at every stage of the evolution from monomer to oligomer.

Proteins that do retain a hint of possibly having passed through a domain-swapped stage are the oligomeric enzymes (Table 3) that have active sites formed from residues on two subunits. If it is assumed that these oligomers evolved from active monomers, the primordial monomer form could have had its active site formed between two domains of a monomer. Then domain swapping would have yielded an active oligomer, which later could have evolved to a stable oligomer by accumulation of mutations that strengthened the secondary domain interface.

**Conclusions.** Our comparison of the structures of the DT monomer and dimer has revealed a mode of intersubunit binding, domain swapping, in which one protein domain is exchanged with an identical domain of the other subunit of the dimer. These "entangling alliances" (Thomas Jefferson, inaugural address, March 4, 1801) between polypeptide chains create new proteins with new properties, as illustrated by a domain-swapped, cooperative form of BS RNase (22). Other intertwined protein oligomers, such as IL-5, could have evolved by domain swapping. We propose in our domain-swapping hypothesis the general features of a pathway for evolution of present-day oligomers by domain swapping. The essential feature of the hypothesis is that at every step of the evolution of the oligomer from the monomer, there is an extensive, tightly binding interface between the monomers of the oligomer. Other models for evolution of oligomers from monomers must postulate a gradual accumulation of comple-