Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 11591–11596, October 1996
Biochemistry

Random circular permutation of genes and expressed polypeptide chains: Application of the method to the catalytic chains of aspartate transcarbamoylase

(protein folding, assembly, and stability/permuted N and C termini/redundant and deleted residues)

RONEY GRAF AND H. K. SCHACHMAN*

Department of Molecular and Cell Biology and Virus Laboratory, University of California, Berkeley, CA 94720-3206

Contributed by H. K. Schachman, August 8, 1996

ABSTRACT **Recent studies on proteins whose N and C termini are in close proximity have demonstrated that folding of polypeptide chains and assembly of oligomers can be accomplished with circularly permuted chains. As yet no methodical study has been conducted to determine how extensively new termini can be introduced and where such termini cannot be tolerated. We have devised a procedure to generate random circular permutations of the catalytic chains of *Escherichia coli* aspartate transcarbamoylase (ATCase; EC 2.1.3.2) and to select clones that produce active or stable holoenzyme containing permuted chains. A tandem gene construct was made, based on the desired linkage between amino acid residues in the C- and N-terminal regions of the polypeptide chain, and this DNA was treated with a suitable restriction enzyme to yield a fragment containing the rearranged coding sequence for the chain. Circularization achieved with DNA ligase, followed by linearization at random with DNase I, and incorporation of the linearized, repaired, blunt-ended, rearranged genes into a suitable plasmid permitted the expression of randomly permuted polypeptide chains. The plasmid with appropriate stop codons also contained *pyrI*, the gene encoding the regulatory chain of ATCase. Colonies expressing detectable amounts of ATCase-like molecules containing permuted catalytic chains were identified by an immunoblot technique or by their ability to grow in the absence of pyrimidines in the growth medium. Sequencing of positive clones revealed a variety of novel circular permutations. Some had N and C termini within helices of the wild-type enzyme as well as deletions and insertions. Permutations were concentrated in the C-terminal domain and only few were detected in the N-terminal domain. The technique, which is adaptable generally to proteins whose N and C termini are near each other, can be of value in relating *in vivo* folding of nascent, growing polypeptide chains to *in vitro* renaturation of complete chains and determining the role of protein sequence in folding kinetics.**

Following the successful *in vitro* construction of a circularly permuted protein, bovine pancreatic trypsin inhibitor (1), other workers have employed this type of polypeptide rearrangement to study structure–function relationships and protein folding. Being formally the equivalent of circularization of a polypeptide chain followed by a cleavage at a site different from the original termini, a circular permutation places charged chain termini at new locations in a protein. Besides disturbing the structure in the region of the cleavage and at the joint between previously existing N and C termini, circular permutation affects a protein in a fundamental way. The continuity of the polypeptide chain is altered. Structures adjacent in the folded protein are moved apart at the sequence level, whereas others are brought together. In terms of possible *in vivo* folding reactions during the growth of the nascent chains, circularly permuted proteins are drastically different from wild-type chains. Since many permuted proteins fold into stable, functional conformations, often both *in vitro* and *in vivo*, the question has arisen whether they follow the same folding pathways as unmutated enzymes. In fact, circularly permuted proteins are becoming an increasingly popular tool in studies relating protein sequence to folding.

Trypsin inhibitor was circularly permuted by chemical linkage of the wild-type N and C termini and subsequent proteolytic cleavage of the protein in a surface loop. Later, molecular genetic approaches via the construction of genes with permuted coding sequences were successfully applied to a variety of proteins: phosphoribosyl anthranilate isomerase (2, 3), mouse dihydrofolate reductase (4), gp120 of human immunodeficiency virus 1 (HIV-1) (5), interleukin 1β (6), mouse ornithine decarboxylase (7), bacteriophage T4 lysozyme (8, 9), the catalytic subunit of *Escherichia coli* aspartate transcarbamoylase (10, 11), ribonuclease T1 (12, 13), *Bacillus* β-glucanase (14), *E. coli* DHFR (15), interleukin 4 (16–18), the SH3 domain of α-spectrin (19), *E. coli* outer membrane protein A (20), *Bacillus stearothermophilus* glyceraldyhyde-3-phosphate dehydrogenase (21), and γB-crystallin (22).

Most of these enzymes were permuted at one or two positions, and only few were subjected to a systematic permutation analysis (11, 13, 19). Although evidence has accumulated that new N and C termini in flexible solvent-exposed loops are readily tolerated, the question arises whether free termini can be moved into other areas of a protein; e.g., within secondary structure elements or into the hydrophobic core. However, even with the availabilty of efficient strategies based on tandem gene constructs that are used to produce "site-directed" circular permutants (6, 11, 12), a one-by-one exhaustive permutation analysis of a protein of several hundred residues in length is hardly a practicable approach. Therefore, we devised a technique to construct a random collection of circularly permuted DNA molecules containing the complete coding sequence for the desired polypeptide chain along with a linker between the original C- and N- terminal regions. These rearranged genes were incorporated into a suitable plasmid for expression of the randomly permuted polypeptide chains in *E. coli*. Colonies containing stable and/or active protein formed from these chains were identified by either an immunoblot assay or by the suppression of auxotrophy because of the presence of active enzyme.

For these studies we have utilized the catalytic (c) polypeptide chain of *E. coli* aspartate transcarbamoylase (ATCase;

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.

Abbreviations: ATCase, aspartate transcarbamoylase; c chain, catalytic polypeptide chain; C trimer, catalytic trimer or subunit; r subunit, regulatory subunit or dimer.

*To whom correspondence and reprint requests should be addressed at: University of California, Department of Molecular and Cell Biology, 229 Stanley Hall # 3206, Berkeley, CA 94720-3206.

aspartate carbamoyltransferase, carbamoyl phosphate:L-aspartate carbamoyltransferase, EC 2.1.3.2) encoded by the *pyrB* gene. ATCase is by far the most highly organized enzyme assembled thus far from circularly permuted chains. The c chain forms two domains, each having a core of a parallel β-sheet packed between an array of α-helices. This polypeptide chain tolerates a variety of circular permutations without losing the ability to assemble into catalytic (C) trimers and the dodecameric holoenzyme that retain activity and even cooperativity (10, 11). This system serves as a useful model for testing a technique to generate randomly circularly permuted chains and to investigate possible rules for inserting termini in various regions of the three-dimensional structures of proteins.

Rationale for Isolation of Protein Containing Randomly Circularly Permuted Chains

The strategy to form and isolate ATCase or other proteins containing randomly circularly permuted c chains involved a series of steps. (*i*) Based on the knowledge of the structure, a decision was made regarding the linkage between the C- and N-terminal regions of the polypeptide chains, and tandem genes were constructed with the appropriate modification to encode the linker between residues near the original termini. For the application described here, this entailed a polypeptide insert which required that the appropriate oligonucleotide be incorporated into the tandem gene construct. (*ii*) This construct was treated with a suitable restriction enzyme to cut each gene at a unique site to yield a rearranged gene fragment that encodes the complete polypeptide chain and linker. (*iii*) The DNA fragment was circularized by the action of DNA ligase. (*iv*) The circular DNA molecules were cleaved at random by a nuclease to yield a collection of linear, permuted genes. (*v*) These permuted DNA fragments, after repairing most nicks and gaps and flushing of ends, were incorporated into a suitable plasmid containing appropriate promoter and termination sequences. This population of plasmids containing the inserts was used for expression of the various circularly permuted polypeptide chains in *E. coli*. (*vi*) Clones expressing the protein molecules were screened by two independent procedures based on either their suppression of auxotrophy because of the presence of active enzyme or their ability to react with antibody to wild-type protein.

Earlier studies (10, 11) have shown that the C- and N-terminal regions of the c chain of ATCase can be linked successfully either by insertion of a linker of 6 residues to bridge the 14-Å distance between residues 310 and 1 or by truncation of the polypeptide chain at residue 306 that is only about 5 Å from residue 1. For the experiments described herein, we departed slightly from the procedure outlined above because of the availability of two previously constructed circularly permuted *pyrB* genes (ref. 11 and Y. R. Yang and H.K.S., unpublished results). One encoded a polypeptide chain starting with an N terminus at position 281 and continuing through residue 310, then through a linker of 6 residues (SGELDM), and finally through residues 1–280. The other gene encoded an analogous chain except that the N terminus was at residue 131 and the chain end was at position 130. Combining these two permuted *pyrB* genes yielded the redundant gene containing codons 131–310, a linker of six codons, and codons 1–280. This construct contained a duplicated portion, codons 131–280, harboring a unique *Bst*EII restriction site. Cutting at these two sites with *Bst*EII yielded a suitable DNA fragment of 316 codons, encompassing one permuted equivalent of the *pyrB* coding sequence. This fragment had single-strand overhangs at each end that facilitated self-ligation to form both closed circular and open circular (relaxed) DNA.

Treatment of the circular forms of the DNA containing the 316 codons with DNase I led to a collection of linear permuted molecules with nicks and gaps that were largely repaired by the use of T4 DNA polymerase and T4 DNA ligase. This procedure, illustrated in Fig. 1, yielded numerous blunt-ended, continuous, double-stranded, permuted *pyrB* genes encoding polypeptides with various N and C termini. These permuted DNA molecules were then incorporated into a plasmid for efficient expression and screening. The plasmid had a strong promoter and allowed precise ligation of blunt-ended fragments between translational start and stop signals. In addition it contained *pyrI*, which encodes regulatory (r) chains of ATCase, so that the coexpressed competent permuted c chains and the r chains could assemble into the dodecameric holoenzyme complex.

We assumed that some circularly permuted c chains and wild-type r chains would assemble into ATCase-like molecules of various degrees of stability and activity. Hence, two different screening procedures were utilized. Clones expressing ATCase-like molecules in sufficient amounts, regardless of their enzyme activity, were detected by a colony-immunoblot technique. Alternatively, colonies containing enzyme activity were identified by their ability to grow on minimal medium in the absence of pyrimidines.

Routine analysis of positive clones at the DNA level involved verification of insert size and approximate location of the permutation site by restriction analysis. At the protein level, expression and activity were evaluated qualitatively by gel electrophoresis of crude extracts to detect native proteins of the size of ATCase using appropriate staining procedures. Selected clones were then analyzed further by DNA sequencing.

Experimental Procedures

Random Permutation of the *pyrB* Coding Sequence. Subsequent to formation of the redundant gene described above, a fragment containing the permuted *pyrB* coding sequence was produced by *Bst*EII digestion (Fig. 1). Circularization of the *Bst*EII fragment was performed with T4 DNA ligase under standard conditions at 16°C for 20 h at a DNA concentration

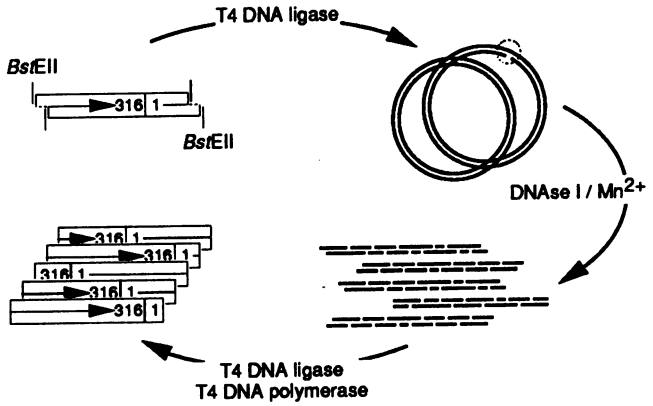

FIG. 1. Method for random circular permutation of the *pyrB* coding sequence. A permuted linear fragment (*Upper Left*) encompassing one equivalent of the coding sequence (948 bp) was isolated from a tandem construct by restriction with *Bst*EII. The tandem construct included codons 131–310, a linker of six codons, and codons 1–280. The linear fragment with overhangs at each end was circularized through the action of T4 DNA ligase to give predominantly closed and relaxed circular DNA (*Upper Right*) that was then treated with DNase I in the presence of Mn^{2+}. The resulting randomly cut linear molecules (*Lower Right*) had nicks, gaps, and staggered ends resulting from the single-strand endonucleolytic digestion. These gaps, nicks, and overhanging ends were largely repaired by the action of T4 DNA polymerase and T4 DNA ligase to yield a library of linear randomly permuted DNA molecules containing the complete coding sequence (*Lower Left*).

of 0.3 mg/ml. More than 80% of the DNA was recovered as circular monomers either in closed or relaxed form, with a background of unligated monomers and linear or circular oligomers.

Random relinearization of the circularized *pyrB* coding sequence was performed by limited digestion with DNase I in the presence of Mn^{2+} (23). DNase I (Pharmacia, 10,000 units/ml) was diluted 1:2000 in the buffer in which it was shipped by the supplier and stored as such for several months without any observed loss of activity. Digests were performed in 50 mM Tris·HCl, pH 7.5/1 mM $MnCl_2$/DNA (100 μg/ml) for 15 min at 16°C, and the reaction was stopped by phenol extraction and ethanol precipitation. Enzyme concentrations were about 1 unit/mg of DNA. Conditions were considered optimal when analytical gel electrophoresis showed that the closed circular species disappeared nearly completely, a smear of degraded DNA was seen in the background, but no oligonucleotides of low molecular weight accumulated at the bottom of the gel.

To close nicks and gaps and flush the ends for subsequent blunt-end cloning, the precipitated product of the DNase I digest was subjected to a repair reaction for 1 h at room temperature utilizing both T4 DNA polymerase and T4 DNA ligase in 16 mM ammonium sulfate/67 mM Tris·HCl, pH 8.8/6.7 mM $MgCl_2$/10 mM 2-mercaptoethanol/10 mM dithiothreitol/6.7 mM EDTA/all four dNTPs (each at 0.13 mM)/1 mM ATP. After this treatment, the linearized species were separated on an agarose gel and recovered.

Expression and Screening. Plasmid pRCP000$^+$ was constructed by the following procedure. In the vector pT7-7 (24) most of the multiple cloning site was replaced by a sequence containing *Nco*I and *Bsu*36I restriction sites. The use of these two enzymes, followed by filling in 5′ overhangs with the Klenow fragment of DNA polymerase, yielded blunt ends with the sequences . . . ATG and TAA . . . , respectively. Two additional stop codons in the second and third reading frame were included to accommodate inserts that were not a precise multiple of 3 bp in length. A copy of the *pyrI* gene was placed downstream from the cloning site to allow polycistronic coexpression of the permuted c chain along with the r chain.

Different promoters for expression of the inserted coding sequence were tested: the *T7* promoter already present in pT7-7, the natural promoter of the *pyrBI* operon, and the isopropyl β-D-thiogalactoside (IPTG)-inducible *tac* promoter. The latter was selected for best compatibility with the two applied screening methods. It allowed sufficiently high expression both on complete and minimal medium even in absence of the inducer IPTG, probably due to insufficient production of lac repressor in the *pyrB*$^-$ strain.

Test ligations with a fragment encompassing the full-length wild-type *pyrB* coding sequence (930 bp) were transformed into the *pyrB*$^-$ strain and reproducibly yielded more than 40% Pyr$^+$ colonies.

The random circularly permuted DNA molecules were cloned into the blunt-ended and dephosphorylated vector pRCP000$^+$ by using approximately equimolar amounts of each DNA which were ligated by T4 DNA ligase in standard buffer at 16°C overnight, followed by transformation into the *pyrB*$^-$ strain, and plating on complete medium containing ampicillin.

Two independent screening methods were applied to identify colonies expressing circularly permuted c chains. A colony immunoblot (25, 26) was used to find clones producing ATCase-like material. Colonies were blotted on nitrocellulose filters and lysed on the membrane by treatment with base. A polyclonal rabbit antibody against purified C subunit was used to detect protein containing permuted c chains; it was visualized by a color reaction using a second antibody-alkaline phosphatase conjugate as described by the supplier (Bio-Rad). Positive colonies were rescued from the original plate and subjected to further analysis.

The requirement for ATCase for cell growth in the absence of pyrimidines allowed a screen for enzyme activity by metabolic complementation. After regeneration from the nitrocellulose blotting procedure, replicas of the original complete medium plates were made on M9 minimal plates (26) containing ampicillin and incubated at 37°C overnight. Due to the *pyrB*$^-$ background, only cells harboring a plasmid encoding sufficient ATCase activity grew, and such clones were selected for analysis.

Clones that were positive in either screen were subjected to a number of preliminary characterizations. Plasmids of interest were recovered, and the size of their inserts and approximate sites of permutation were determined by restriction analysis. Crude cell extracts were prepared from small (1–2 ml) cultures and separated on native polyacrylamide gels. Overexpressed ATCase holoenzyme was identified as a prominent high molecular weight band after staining with Coomassie brilliant blue. Alternatively, the use of a sensitive activity stain (27) allowed detection of trace amounts of activity in the extract. Combined results from both methods gave qualitative estimates on the amount of stable, soluble ATCase in the extract and its activity. Based on these preliminary analyses of DNA and protein, clones were chosen for sequence determination (28) to determine the precise site of the permutation.

Preliminary Characterization of ATCase-Like Molecules Containing Circularly Permuted c Chains. Purification of the holoenzyme was facilitated by incorporating a hexahistidine affinity tag onto the N terminus of the r chain of ATCase. This was achieved by inserting an oligonucleotide containing five histidine codons into the *pyrI* gene between start codon Met-1 and codon His-3 and eliminating codon Thr-2. In plasmid pRCP000$^+$ and its derivatives, containing circularly permuted variants of *pyrB*, the *pyrI* segment was replaced by this modified gene to encode the r chain with the hexahistidine tag. After coexpression in *E. coli*, ATCase-like molecules containing tagged r chains were purified from crude extracts by affinity chromatography on a metal-chelating column (Ni^{2+}-NTA agarose, Qiagen, Chatsworth, CA). Protein was bound to the column in 50 mM Tris·HCl at pH 8.0 containing 200 mM KCl, and eluted with 50 mM sodium phosphate at pH 6.5 containing 200 mM KCl with a gradient of 0–500 mM imidazole. About 15–40 mg of purified holoenzyme were isolated from a 1-liter culture.

Enzyme assays and sedimentation velocity measurements were performed by methods used previously in this laboratory (10, 11). Molecular masses of polypeptide chains were kindly determined by David King (Howard Hughes Medical Institute, University of California, Berkeley). The c and r chains were separated on a Vydac C_{18} reverse-phase HPLC column, and masses were measured on a Hewlett–Packard S989A electrospray ionization mass spectrometer.

Results

Formation of a Random Collection of Linear Permuted DNA Molecules Containing the Coding Sequence for c Chains. The gel electrophoresis patterns in Fig. 2 illustrate the effectiveness of the individual enzymatic reactions used to prepare the permuted DNA molecules. Lane 2 shows the single band representing the DNA fragment of 316 codons (948 bp) excised from the two linked circularly permuted *pyrB* genes. This linear fragment with single-strand overhangs contains the complete coding sequence of the c chain with the N terminus at residue 162, and it includes the nucleotide sequence for the 6-amino acid linker between residues 310 and 1. Self-ligation of the ends of the fragment through the action of DNA ligase produced predominantly closed and relaxed circular DNA as shown by the pattern in lane 3. Some linear molecules and aggregates are present as seen in the gel pattern. However, there was no need to remove this small amount of unligated

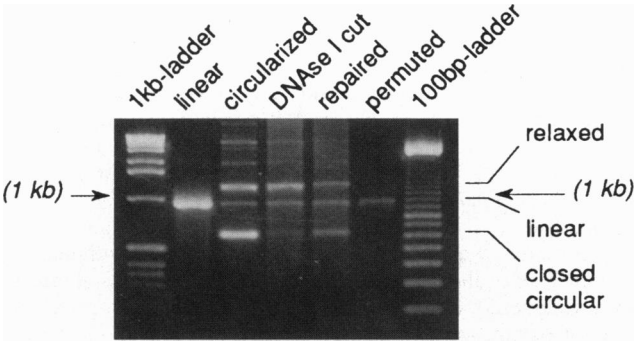

FIG. 2. Formation of a collection of randomly permuted DNA molecules containing the complete coding sequence for c chains of ATCase. All samples were subjected to electrophoresis in a 1.5% agarose gel in a buffer containing 100 mM Tris·HCl, 120 mM sodium borate (pH 8.4), 0.1 mM EDTA, and ethidium bromide (0.5 μg/ml). Lanes 1 and 7 show molecular weight markers obtained from GIBCO/BRL and Pharmacia, respectively. Lane 2 shows the pattern for the DNA fragment (948 bp) containing the complete coding sequence for the c chain. Circularization of this fragment by self-ligation of the single strand overhangs gave the preparation shown by the pattern in lane 3. Digestion of this preparation with DNasc I to give a random collection of linear fragments yielded a sample exhibiting the pattern in lane 4. Treatment of these DNA molecules to repair nicks, gaps, and overhangs by incubating with T4 DNA polymerase and T4 DNA ligase yielded the pattern in lane 5, followed by fractionation to give the linear permuted double-stranded DNA molecule shown by the pattern in lane 6.

linear molecules because the subsequent random endonucleolytic cleavage at a frequency of one cut per molecule would degrade most of it. In addition, filling in the 5′ overhangs of the *Bst*EII fragment by polymerase does not yield an in-frame translatable unit, and therefore, the linear fragments do not produce an unwanted background. The collection of mostly circular DNA molecules was subjected to DNase I under conditions aimed at producing a double-stranded break in each circular molecule only once. As seen in the pattern in lane 4, there are bands representing various species including one corresponding to molecules about 1 kb in size. Repair of the nicks by T4 DNA polymerase and T4 DNA ligase had only a slight impact on the distribution of species on an agarose gel (lane 5) but was crucial for the successful recovery of the functional linearized DNA. This series of steps produced largely blunt-ended permuted DNA molecules (lane 6) containing the coding sequence of the c chain of ATCase with an overall yield of about 5% of the initial material. The transformation frequency with the final collection of permuted DNA molecules in the cloning experiments was comparable to that in control experiments with other blunt-ended DNA fragments.

***In Vivo* Formation of Random ATCase-Like Molecules Containing Circularly Permuted c Chains.** Analysis of colonies by the immunoblot assay showed positive clones with a frequency of 1–2%, and pyrimidine prototrophs were observed at a 2–4% frequency. Colonies fulfilling both conditions occurred at a frequency of 0.5–1.0%. These frequencies seem reasonable since some colonies would be expected to have no insert, another group (50%) would probably have the permuted DNA insert in the incorrect orientation, and 33% of those with the proper orientation would be expected to carry the permuted insert with its 5′ end in the correct reading frame. Hence a maximum probability of only 10–15% could be expected for a given colony to express circularly permuted c chains. Even this is likely to be an overestimate since additional factors, such as multiple cuts by DNase I and gaps or redundant ends resulting from flushing of staggered ends generated by DNase I, would lead to a lower probability of positive clones. Of greatest significance in lowering the yield of positive colonies is the likelihood that some permutation events in unfavorable regions of the structure would lead to chains that cannot fold properly and assemble into ATCase-like molecules. In view of these considerations, the overall procedure is reasonably efficient.

Restriction analysis of plasmids recovered from positive clones showed that the size of inserts was usually close to the expected 948 bp. Clones with inserts too short by 50–100 bp occurred at a frequency of about 15% and were not analyzed further. The site of permutation could often be judged within 10–15 amino acid residues. Restriction mapping indicated extensive heterogeneity in the population of clones, suggesting that no particular sites were heavily preferred by DNase I in the random cutting of the circular DNA. About 80% of the permutation events, however, occurred in the last third of the sequence. Strikingly, only 10 of 83 clones analyzed by restriction were permuted in the N-terminal (carbamoyl phospate-binding domain) of the polypeptide chain.

Cell extracts of positive clones were analyzed on native polyacrylamide gels to determine the level of expression and electrophoretic behavior of permuted ATCases. Only extracts of strains that gave a positive signal in the immunoblot showed the characteristic holoenzyme band when stained with Coomassie blue, indicating that the colony-immunoblot technique detected ATCase only when expression was reasonably high. In contrast, when gels were examined by the highly sensitive ATCase activity stain (27), no band was detected in many strains that were able to grow without supplemented pyrimidines. We conclude that the complementation assay could detect clones with only trace amounts of active ATCase or enzyme with extremely low activity. In no case did we detect bands indicating a permuted species forming active C trimers that were unable to assemble into holoenzyme.

Location and Identification of N and C Termini in Circularly Permuted c Chains of ATCase. The results of the DNA sequence determinations of 18 different permuted clones are summarized in Fig. 3. It also shows in circular form a topological map of the c chain in terms of β-strands, α-helices, and loops based on the crystallographic structure of wild-type ATCase (29, 30). As expected, the size of the chains in the various permutants varied slightly due to correction of the staggered ends produced by the action of DNase I. The maximum deviation from the wild-type sequence of 948 bp was a gain of 14 nt or a loss of 27 nt. Since many inserts were not an integral number of codons in length, they had to utilize one of the "rescue" stop codons provided by the plasmid in the second or third reading frame. The resulting polypeptide chain thus contained either two or four additional amino acids and had as C-terminal regions either (L/I/V)-R or XKVN. Based on sequences of the different DNAs, the encoded permuted polypeptide chains varied in length from 307 to 324 amino acid residues.†

Most of the N termini in the circularly permuted c chains occured in loop regions, sometimes just before or after a helix or β-strand as in A32/A32NKVN, Q196/P195HKVN, A208/W209, and Y285/A283. However, seven permutants had N termini within α-helices. Two species started at Gln-137, which is located within helix 5. These permutants differ at their C termini, with one, Q137/S131, lacking five residues including two of helix 5 and three of the preceding loop. The other,

†Six additional positive permutants, not included in Fig. 3, were identified by the enzyme activity stain. DNA sequences showed that the inserts were not in frame with the ATG start codon. All of these rearranged DNAs carried a methionine codon located 11–23 nt after the vector-insert junction. Presumably these downstream ATG codons served as alternative translation initiation sites. Three permutants are of potential interest since the DNA sequences indicated deletions of about 15 amino acid residues from the polypeptide chain. As yet purified proteins encoded by these permuted DNAs are not available.

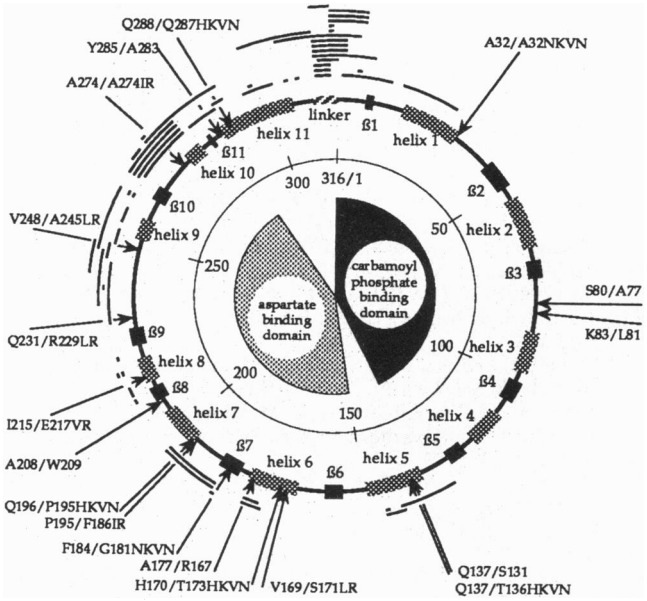

FIG. 3. Circular topological map of the c chain in ATCase along with location and identification of N and C termini in circularly permuted chains. Secondary structural elements, β-strands and α-helices are designated along the circle as boxes, solid and stippled, respectively. The carbamoyl phosphate binding domain contains residues 1–134 and the aspartate binding domain includes residues 150–284. Arcs on the outside of the circle were the results of mapping by restriction analysis, and they designate random permutants identified by the immunoblot assay or by suppression of auxotrophy due to the presence of active ATCase in the colonies. A permuted site for the new termini is located within or close to each arc with an accuracy of 10–15 amino acid residues. Permutants for which complete DNA sequences were obtained are designated by arrows; the first (N terminal) residue is indicated by the one letter code and the number of the residue based on the wild-type sequence. This is followed by a slash (/) indicating the bulk of the DNA sequence followed by the determined C-terminal region, either a single amino acid residue with the numbered positon or additional residues resulting from the use of stop codons in the second or third reading frame. The topological map is based on the crystallographic structure (29, 30) and coordinates deposited in the Brookhaven Protein Data Bank (registry number 6ATI).

Q137/T136HKVN, contained four additional residues. Both proteins were stable with Q137/S131 virtually devoid of enzyme activity and Q137/T136HKVN exhibiting moderate activity. Other permutants, H169/S171LR and H170/T173HKVN, also had N termini within an α-helix of the wild-type enzyme. These enzymes with termini in helix 6 were stable but inactive. In contrast, permutant I215/E217VR with an N terminus in helix 8 was active as well as stable. One permutant, Q288/Q287HKVN, with an N terminus in helix 11, is of particular interest since previous studies (31, 32) indicated that few substitutions in that helix could be tolerated. In only one permutant, F184/G181NKVN, was an N terminus found within a β-strand of wild-type ATCase.

Inferences about protein structures based solely on DNA sequences may be misleading because of possible intracellular processes such as recombination, post-translational modification, and proteolysis. Therefore, we purified seven proteins for preliminary characterization. Five proteins had c chains with molecular masses, determined by electrospray ionization mass spectrometry, in excellent agreement with the values calculated from the DNA sequences. On the basis of results from mass spectrometry, we conclude that three of the permuted c chains, such as Q137/S131, had N-terminal methionines. Permutant A32/A32NKVN had no N-terminal methionine; and one, H169/S171IR, exhibited partial processing with both species detected by mass spectrometry.

Two purified proteins are of particular interest in illustrating the utility of producing randomly permuted polypeptide chains. The protein A32/A32NKVN, in which the new N terminus is just at the end of helix 1, contains a 5-amino acid insert at its C terminus. The holoenzyme of moderate activity has a sedimentation coefficient virtually identical to that of wild-type ATCase. In contrast, Q137/S131 has its N terminus within helix 5 and five amino acids were deleted. This holoenzyme, though exhibiting very little enzyme activity, is stable despite the introduction of an N terminus in a region that is helical in the wild-type enzyme.

Discussion

The results summarized in Fig. 3 demonstrate the efficiency and utility of the technique described herein for producing a random collection of protein molecules containing circularly permuted polypeptide chains. Although the method was tested only with c chains of ATCase, the individual steps are generally applicable to other proteins whose N and C termini are located in reasonably close proximity. After a decision is made about the linkage between the wild-type terminal regions, either through a peptide insert or a deletion, and the appropriate tandem gene arrangement is constructed, all the subsequent steps to form the circular DNA containing the coding sequence are straightforward. It is in the next step aimed at producing a random collection of linear permuted DNA molecules where more approaches are clearly needed.

Randomness in the permuted coding sequences encompassed by the linear DNAs was achieved with DNase I as a relatively nonspecific endonuclease. Although this enzyme is not totally nonspecific, it has been used in applications requiring random cutting of DNA and was found to have little sequence specificity (33). Our results show that DNase I is an efficient and convenient tool for producing a collection of linear DNAs from closed and relaxed circular forms. The use of DNase I is not without disadvantages, however. Through single-strand cleavages, DNase I causes nicks, gaps, and staggered ends that require repair by treatment with T4 DNA polymerase and T4 DNA ligase. The repair of staggered ends after DNase I cleavage leads to the loss or gain of some nucleotides requiring correction of the reading frame by utilization of alternative stop codons that in turn results in additional amino acid residues in the polypeptide chain. Hence few of the chains produced by the described technique are perfectly permuted. This is demonstrated in Fig. 3 where variability of the C terminus relative to a particular N-terminal site is evident. The redundancies, deletions, and insertions in the polypeptide chains resulting from the DNase I step can be an advantage in pursuing the ultimate goal of understanding where N and C termini can and cannot be tolerated. This procedure producing insertions and deletions may yield functional permutants that may not be viable in the perfectly permuted form.

Other methods to convert circular DNA into a random collection of linear double-stranded molecules should certainly be explored. Endonuclease I, for example, cleaves both strands of a DNA helix, but preparations also contain a significant exonucleolytic component that acts at the site of the original cleavage. If that activity could be removed, endonuclease I would be a good candidate for the purpose outlined here. It is also possible that a mixture of many restriction enzymes of different specificities that cut to yield blunt ends could be used to make linear permuted DNAs from the circular forms. Type I restriction enzymes that bind at a unique site and cleave elsewhere in the DNA at random may also be worthy of consideration. Use of this approach may require making changes in the DNA sequence, perhaps in the linker region, to create the required recognition site. Physical means

such as sonication rather than an endonuclease to produce double-strand breaks may also be of use.

Almost 80% of the permutations were localized in the last third of the sequence of the wild-type chains. Mapping by restriction analysis showed that 15 sites were located in the region of the wild-type termini, i.e., in the loop between helix 11 and β-strand 1. Clustering of permutants in this "wild-type loop" indicates that this area can readily accomodate free N and C termini. In this regard it is of interest that so few termini were found in the N-terminal domain of the protein; three were found in this random process and two others are known from site-directed permutations (11). Is this part of the protein structure much less tolerant to the introduction of new termini? Additional specific permutations in this domain are clearly needed.

The random permutation analysis of the c chain of ATCase has yielded the important result that N and C termini can be placed in regions that were helical in the wild-type enzyme. To what extent is the structure altered by the introduction of these termini? Is the insertion of additional residues as in Q137/T136HKVN helpful or would a deletion as in Q137/S131 lead to greater stability? These and related questions about introduction of termini within β-strands illustrate the need for both detailed structural and functional studies on the purified proteins and expanding the library to include an exhaustive set of all functional permuted c chains. Finding unexpected permutants by having the cell select those altered chains that are competent for assembly into stable functional proteins should prove of value in studies of the folding of polypeptide chains and the stability of proteins.

We thank Ying R. Yang and Shaival Shah for help and valuable suggestion during the course of this work. We thank Dr. David King of the Howard Hughes Medical Institute, University of California, at Berkeley, for determing the molecular weights of the polypeptide chains by mass spectrometry. This research was supported by National Institute of General Medical Sciences research Grant GM 12159 and a fellowship to R.G. from the Swiss National Science Foundation.

1. Goldenberg, D. P. & Creighton, T. E. (1983) *J. Mol. Biol.* **165,** 407–413.
2. Luger, K., Hommel, U., Herold, M., Hofsteenge, J. & Kirschner, K. (1989) *Science* **243,** 206–210.
3. Luger, K., Szadkowski, H. & Kirschner, K. (1990) *Protein Eng.* **3,** 249–258.
4. Buchwalder, A., Szadkowski, H. & Kirschner, K. (1992) *Biochemistry* **31,** 1621–1630.
5. Pollard, S. R., Rosa, M. D., Rosa, J. J. & Wiley, D. C. (1992) *EMBO J.* **11,** 585–591.
6. Horlick, R. A., George, H. J., Cooke, G. M., Tritch, R. J., Newton, R. C., Dwivedi, A., Lischwe, M., Salemme, F. R., Weber, P. C. & Horuk, R. (1992) *Protein Eng.* **5,** 427–431.
7. Li, X. & Coffino, P. (1993) *Mol. Cell. Biol.* **13,** 2377–2383.
8. Zhang, T., Bertelsen, E., Benvegnu, D. & Alber, T. (1993) *Biochemistry* **32,** 12311–12318.
9. Zhang, T., Bertelsen, E. & Alber, T. (1994) *Nat. Struct. Biol.* **1,** 434–438.
10. Yang, Y. R. & Schachman, H. K. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 11980–11984.
11. Zhang, P. & Schachman, H. K. (1996) *Protein Sci.* **5,** 1290–1300.
12. Mullins, L. S., Wesseling, K., Kuo, J. M., Garrett, J. B. & Raushel, F. M. (1994) *J. Am. Chem. Soc.* **116,** 5529–5533.
13. Garrett, J. B., Mullins, L. S. & Raushel, F. M. (1996) *Protein Sci.* **5,** 204–211.
14. Hahn, M., Piotukh, K., Borriss, R. & Heinemann, U. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 10417–10421.
15. Protasova, N. Yu., Kireeva, M. L., Murzina, N. V., Murzin, A. G., Uversky, V. N., Gryaznova, O. I. & Gudkov, A. T. (1994) *Protein Eng.* **7,** 1373–1377.
16. Kreitman, R. J., Puri, R. K. & Pastan, I. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 6889–6893.
17. Kreitman, R. J., Puri, R. K. & Pastan, I. (1995) *Cancer Res.* **55,** 3357–3363.
18. Kreitman, R. J., Puri, R. K., McPhie, P. & Pastan, I. (1995) *Cytokine* **7,** 311–318.
19. Viguera, A. R., Blanco, F. J. & Serrano, L. (1995) *J. Mol. Biol.* **247,** 670–681.
20. Koebnik, R. & Kramer, L. (1995) *J. Mol. Biol.* **250,** 617–626.
21. Vignais, M. L., Corbier, C., Mulliert, G., Branlant, C. & Branlant, G. (1995) *Protein Sci.* **4,** 994-1000.
22. Komar, A. A. & Jaenicke, R. (1995) *FEBS Lett.* **376,** 195–198.
23. Melgar, E. & Goldthwait, D. A. (1968) *J. Biol. Chem.* **243,** 4409–4416.
24. Tabor, S. & Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 1074–1078.
25. Sambrook, L., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed., pp. 12.18–12.24.
26. Laurenti, P., Graba, Y. & Pradel, J. (1993) *Trends Genet.* **9,** 335–336.
27. Bothwell, M. A. (1975) Ph.D. thesis (University of California, Berkeley).
28. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
29. Stephens, R. C., Chook, Y. M., Cho, C. Y., Lipscomb, W. N. & Kantrowitz, E. R. (1991) *Protein Eng.* **4,** 391–408.
30. Lipscomb, W. N. (1994) *Adv. Enzymol.* **68,** 67–151.
31. Peterson, C. B. & Schachman, H. K. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 458–462.
32. Peterson, C. B. & Schachman, H. K. (1992) *J. Biol. Chem.* **267,** 2443–2450.
33. Anderson, S. (1981) *Nucleic Acids Res.* **9,** 3015–3027.