Vol. 169, No. 1, 1990 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
May 31, 1990 Pages 136-142

cDNA CLONING AND SEQUENCING FOR THE IMPORT PRECURSOR OF SUBUNIT B IN H+-ATP SYNTHASE FROM RAT MITOCHONDRIA‡

C. Tsurumi,* Y. Yoshihara,* F. Osaka,* F. Yamada,* I. Tani,* T. Higuti,*#
M. Shimizu,§ K. Oeda,§ H. Ohkawa,§ H. Toda,† T. Kakuno,¶ F. Sakiyama,†
A. Kumatori,⊥ K. Tanaka,⊥ and A. Ichihara⊥

*Faculty of Pharmaceutical Sciences, University of Tokushima, Shomachi, Tokushima 770, Japan

§Biotechnology Laboratory, Takarazuka Research Center, Sumitomo Chemical Company Limited, Takatsukasa, Takarazuka, Hyogo 665, Japan

†Division of Protein Chemistry, ¶Division of Enzymology, Institute for Protein Research, Osaka University, Yamadaoka, 3-2, Suita, Osaka 565, Japan

⊥Institute for Enzyme Research, The University of Tokushima, Tokushima 770, Japan

Received April 14, 1990

Summary: The nucleotide sequence of the import precursor of subunit b of rat liver H+-ATP synthase has been determined from a recombinant cDNA clone isolated by screening a rat liver cDNA library with a probe DNA. The sequence was composed of 1,124 nucleotides including a coding region for the import precursor of subunit b and noncoding regions of both the 5'- and 3'-sides. The import precursor of subunit b and its mature polypeptide deduced from the open reading frame consisted of 256 and 214 amino acid residues with a molecular weight of 28,867 and 24,628, respectively. The presequence of 42 amino acids could be the import signal peptide which serves to direct the protein into the mitochondrial matrix.

H+-ATP synthase is a multi-subunit complex that utilizes a trans-membrane proton gradient ($\Delta\mu H^+$) to form ATP (1). H+-ATP synthase is composed of two domains: a hydrophilic portion called F_1, which is the catalytic site of ATP synthesis, and a membranous domain called F_o, which is responsible for energy transduction (2-4). However, it is still unknown how this enzyme converts $\Delta\mu H^+$ into energy for ATP synthesis (cf. ref. 5-8).

Recently, we developed a new simple method of the purification of H+-ATP synthase from rat liver mitochondria, which had oligomycin-sensitive high ATPase activity (9). Then, we purified subunit b and factor 6 from the purified H+-ATP synthase, which were first found from rat (9).

‡This work was supported in part by Grants-in-Aid for Scientific Research (Nos. 63480502 and 01880022) from the Ministry of Education, Science and Culture of Japan.

#To whom all correspondence should be addressed.

0006-291X/90 $1.50

In the present work, we showed the cloning and sequence of cDNA of the import precursor of the rat subunit b of H^+-ATP synthase. This is the first report on the mitochondrial import precursor of subunit b.

Materials and Methods

Materials. The following reagents were used: [α-^{32}P] dCTP (111 TBq/mmol) and [γ-^{32}P] ATP (185 TBq/mmol) from Amersham Corp.; various restriction endonucleases from Takara Shuzo, Bethesda Research Laboratories, and New England Biolabs.

Purification of H^+-ATP synthase and subunit b. H^+-ATP synthases were purified to homogeneity from rat liver mitochondria by our recently developed method (9). Subunit b was purified directly from the purified H^+-ATP synthase by reverse-phase high performance liquid chromatography on a Asahipak ODP-50 (Asahi Chemical Industry Co., Ltd., 6.0 x 150 mm) (9).

Sequencing of subunit b. The purified subunit b was sequenced in an Applied Biosystems, 470A protein sequencer.

Synthesis of Oligonucleotide Probe. Oligonucleotide hybridization probe was synthesized in an Applied Biosystems, Model 380A synthesizer. 5'-End labeling of oligodeoxyribonucleotide with [γ-^{32}P]ATP and T4 polynucleotide kinase was carried out as described (10).

Preparation of Poly(A)$^+$RNA. Total RNA was extracted from rat liver by the guanidinium thiocyanate method (11), and poly(A)$^+$RNA was isolated by oligo(dT)-cellulose chromatography (12).

Construction of a cDNA Library. A cDNA library of liver from Wistar rats was constructed in a phagemid expression vector, Bluescript KS M13$^+$ (Stratagene). Double-stranded DNA complementary to rat liver poly(A)$^+$RNA was synthesized by the method described by Fujiwara et al.(13), except that 5'-TAGGTCGACGCGGCCGCTTTTTTTTTTTTTTT-3' was used as primer for synthesis of the first strand of cDNA. This primer includes unique restriction endonuclease sites for *Not*I (5'-GCGGCCGC-3') and *Sal*I (5'-GTCGAC-3') at the 5'-side of oligo(dT)$_{15}$. The double-stranded DNA mixtures synthesized were treated with Klenow fragment to blunt their ends. The resulting DNA mixtures were digested with *Not*I. Excess linkers and *Not*I-digested linker fragments were removed by fractionation on a QIAGEN-tip column (Funakoshi Co. Ltd., Tokyo, Japan). The DNA mixtures were then ligated with vector Bluescript. The vector had been digested with *Not*I and *Eco*RV at the multicloning site. *Escherichia coli* HB101 competent cells (Takara Shuzo) were transformed with the ligated DNA.

Screening of the cDNA Library by Oligonucleotide Hybridization. For isolation of cDNAs for subunit b, about 40,000 transformants were screened by hybridization with ^{32}P-labeled oligodeoxyribonucleotide probe synthesized as described above. Colony hybridization was carried out by the method of Hanahan and Meselson (14). Briefly, transformants that had been replicated onto duplicate nylon filters were prehybridized overnight at 50°C in a solution of 50 mM Tris-HCl (pH 7.5), 1 M NaCl, 1% SDS, and 200 μg/mL yeast RNA and hybridized for 5 h in the same solution containing 100 μg/mL sonicated salmon sperm DNA and labeled probe (approximately 10^6 cpm/mL), which had a specific activity of about 6.5 x 10^6 cpm/pmol. The filters were washed sequentially with 6x SSC containing 0.1% SDS at room temperature for 5 min, at 50°C for 10 min. The filters were then exposed to X-ray film at -70°C with an intensifying screen.

DNA Sequencing. DNA sequencing was carried out by the dideoxy chain-termination method (15) with a 7-DEAZA sequencing kit from Takara Shuzo.

Sequence Analysis of DNA and Protein. Analysis of sequence data was carried out using the DNASIS program (Hitachi Software Engineering Co., Ltd.).

Results and Discussion

Sequencing of Subunit b Purified from Rat Mitochondrial H^+-ATP synthase. H^+-ATP synthase was purified from rat liver mitochondria by our recently developed procedure (9). Then, subunit b was isolated from the purified H^+-ATP synthase by a reverse-phase column of Asahipak ODP-50 as described previously (9). The purified subunit b was analyzed with an Applied Biosystems 470A protein sequencer, and the subunit b was found to have the following sequence: PLPPLPEYGGKVRLGLIPEEFFQFLYP.

The subunit b was also cleaved with cyanide bromide, and then the obtained peptide fragments was applied to the reverse-phase column (data not shown). The peptide fragment eluted as the main peak was analyzed with the protein sequencer and the peptide was found to have the following sequence: INWVEKHVIQ-I-AQQEKE (-, not identified).

Isolation of cDNA Clones Encoding Subunit b. A rat cDNA library was constructed with the phagemid Bluescript vector using poly(A)$^+$RNAs extracted from rat liver as described under Materials and Methods. For isolation of cDNA for subunit b, this library was screened by hybridization with synthetic deoxyribonucleotide as probe. The probe DNA was synthesized, based on the determined 20 amino acids sequence from the N-terminus of rat subunit b, the sequence of cDNA for bovine subunit b (16), and the coding frequencies of eukaryotic proteins (17) as follows: 5'-TTCCTCAGGGATCAGCCCCAGACGAACTTTTCCTCCATATTCAGGAAGAGGTGGCA GAGG-3'. We screened about 40,000 colonies of a rat liver cDNA library with the synthesized 60-mer oligonucleotide. Four cDNA clones that gave strongly positive signals for the probe DNA were isolated from the library by colony hybridization techniques. As the cDNAs of all four clones gave similar cleavage maps with several restriction endonucleases, clone with a largest cDNA insert of 1.1-1.2-kb length plus a poly(A) tail, was subjected to cDNA sequence determination.

Primary Structure Deduced from the cDNA Sequence. The nucleotide sequence of the rat subunit b cDNA was determined by the strategy illustrated in

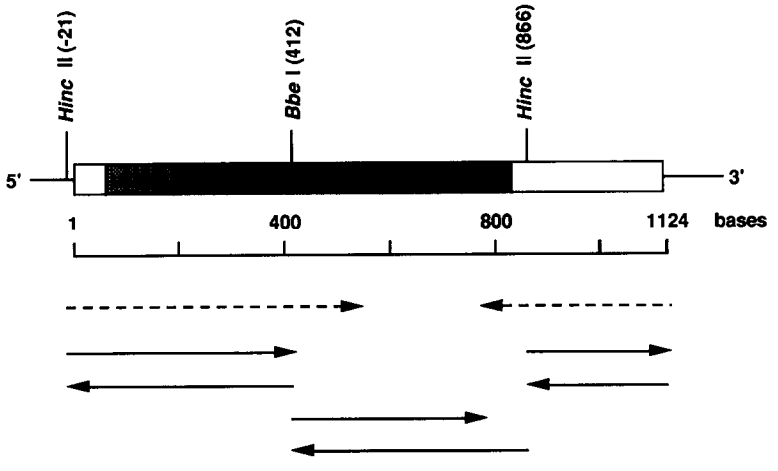

Fig. 1. Restriction endonuclease map of cloned cDNA for rat subunit b of H$^+$-ATP synthase and the sequencing strategy. The half-tone, solid, and open boxes show the coding regions for presequence and mature subunit b, and 5'- and 3'-noncoding regions, respectively. The solid lines indicate the sequence of the vector Bluescript KS$^+$. Sequenced regions are shown by horizontal arrows. The sequences of the 5'-end and 3'-end of cDNA (dashed arrow) were determined by direct sequencing of an isolated clone with the primers of T_3 and T_7 promotor of Bluescript KS$^+$, respectively. The sequences of other regions (solid arrows) were determined by sequencing of subclones with the primers of KS and SK promotor of Bluescript KS$^+$.

```
   1  GA GTG TCT CGG GTC GGG GTC ACA GGG ACA TCG GTA CCG CTG TCT GGA TCT TCG   53

  54  CAG ACA ATG CTG TCC CGG GTG GTA CTT TCT GCC GCC GCC ACA GCG GCC CCG TGT  107
   1          M   L   S   R   V   V   L   S   A   A   A   T   A   A   P   C   16
              └──►Import Signal Peptide
 108  CTG AAG AAC GCG GCC GTC CTG GGT CCA GGG GTT TTA CAG GCA ACA AGG GTC TTT  161
  17  L   K   N   A   A   V   L   G   P   G   V   L   Q   A   T   R   V   F   34

 162  CAC ACA GGA CAG CCT CGC CTT GCC CCT CTA CCA CCT CTT CCT GAA TAT GGA GGA  215
  35  H   T   G   Q   P   R   L   A │P   L   P   P   L   P   E   Y   G   G   52
                                     └──►Mature subunit b
 216  AAA GTA CGT CTT GGG CTG ATT CCT GAG GAA TTT TTC CAG TTC CTT TAC CCT AAG  269
  53  K   V   R   L   G   L   I   P   E   E   F   F   Q   F   L   Y   P   K   70

 270  ACT GGT GTA ACA GGA CCC TAC GTG CTT GGA ACT GGG CTT AGC TTG TAC TTT CTA  323
  71  T   G   V   T   G   P   Y   V   L   G   T   G   L   S   L   Y   F   L   88

 324  TCC AAA GAA ATA TAT GTG ATT ACC CCA GAG ACC TTC TCT ACC ATA TCA GTA GTA  377
  89  S   K   E   I   Y   V   I   T   P   E   T   F   S   T   I   S   V   V  106

 378  GGG TTG ATT GTC TAT GTG ATT AAG AAA TAT GGC GCC TCT ATT GGA GAA TTT ATT  431
 107  G   L   I   V   Y   V   I   K   K   Y   G   A   S   I   G   E   F   I  124

 432  GAC AAA CTT AAT GAG GAA AAA ATT GCT CAA CTA GAA GAA ATA AAG CAG TCA AGC  485
 125  D   K   L   N   E   E   K   I   A   Q   L   E   E   I   K   Q   S   S  142

 486  ATG AAG CAA ATC CAG GAT GCA ATC AAC AGG GAG AAG GCA CAG CAG GCA CTG GTT  539
 143  M   K   Q   I   Q   D   A   I   N   R   E   K   A   Q   Q   A   L   V  160

 540  CAG AAG CGC CAC TAC CTC TTC GAT GTT CAG AGG AAT AAC ATT GCC CTG GCC TTG  593
 161  Q   K   R   H   Y   L   F   D   V   Q   R   N   N   I   A   L   A   L  178

 594  GAG GTC ACT TAC CGG GAA CGA CTA CAT AAA GCA TAT AAG GAG GTA AAG AAT CGC  647
 179  E   V   T   Y   R   E   R   L   H   K   A   Y   K   E   V   K   N   R  196

 648  CTG GAC TAC CAT ATT TCT GTA CAG GAC ATG ATG CGT CGC AAG GAG GGA GAG CAC  701
 197  L   D   Y   H   I   S   V   Q   D   M   M   R   R   K   E   G   E   H  214

 702  ATG ATA AAC TGG GTG GAG AAG CAT GTG ATA CAG AGC ATT TCT GCA CAG CAG GAA  755
 215  M   I   N   W   V   E   K   H   V   I   Q   S   I   S   A   Q   Q   E  232

 756  AAG GAG ACC ATT GCC AAG TGC ATC GGA GAT CTA AAG ATG CTT GCA AAG AAG GCT  809
 233  K   E   T   I   A   K   C   I   G   D   L   K   M   L   A   K   K   A  250

 810  CAA GCT CAG CCA ATT ATG TGA ATG TGT CTA TCT CAG TTG TGA TAG CCA GAG AGG  863
 251  Q   A   Q   P   I   M   *                                              256

 864  GTT AAC TTA AAT GAG AAC TAC TCT ACT TGA AGA ACT CTT TCT GTA TTG CTG TCT  917

 918  AAT GAA ATT GTG GTG TAC CTT TCC TAA GAA ATG ATC AAG GTC CAT TTA GTG GCT  971

 972  TAA CCC TAT TTC TGA TCC TTG CTC TGT ATT CGG AGT TGT CTG ATC ACA TTT GAG 1025

1026  TAA GCA ATT TGC AGC AAC TCG CTG CCC AAC AGA AAT TAC CAA GTT ATA GTT TAA 1079

1080  ACT TGT AAT TAG TTA AAA CAT CTT ACA ATA AAA TGT TTG AAA CAG            1124
```

Fig. 2. Nucleotide sequence of the cDNA insert encoding the possible import signal peptide and subunit b of rat liver H^+-ATP synthase and the amino acid sequence deduced from its open reading frame. Nucleotides are numbered in the 5'- to 3'-direction. The predicted amino acid sequences of subunit b and its possible import signal peptide are shown below the nucleotide sequence. Amino acid residues are numbered from the N-terminus. A solid lines show the amino acid sequences determined by Edman degradation. Amino acids shown by dotted lines in the sequence was not identified by chemical analysis. The possible polyadenylation signal AATAAA is doubly underlined.

Figure 1. The resulting nucleotide sequence of the cDNA and the primary structure of the mature subunit b and also the possible import signal peptide deduced from the cDNA sequence are shown in Figure 2. The sequence of 1,124 nucleotides included the entire coding region and 5'- and 3'-noncoding regions.

The 3'-noncoding region consisted of 297 nucleotides, excluding the poly(A) tail. A polyadenylation signal (AATAAA) that is common in eukaryotic mRNAs (18,19) was located 14 nucleotides upstream from the poly(A) addition site. Part of the poly(A) sequence (approximately 15 nucleotides) was seen at the 3'-end of this cDNA clone.

The mature protein coding region is 642 nucleotides long, is capable of encoding 214 amino acids with a molecular weight of 24,628, and extends from CCT at nucleotide position 186-188 to nucleotide position 827. This coding sequence is followed by the termination codon TGA. The partial amino acid sequences determined by automated Edman degradation were found to be in complete accordance with those deduced from the cDNA sequence. The amino acid sequence of rat subunit b deduced from the cDNA was highly homologous with that of bovine subunit b (16) (matching percentage, 84%).

The possible coding region for the import signal peptide is seen at the upstreem of the mature protein coding region from only one ATG codon in the same frame at nucleotide position 60-62 to nucleotide position 185, which is capable of encoding 42 amino acids. The formal possibility exists that the coding sequence for the presequence extends beyond the end of the sequence determined in the cDNA. However, this possibility could be excluded by the following reason.

The import signal peptides of mitochondrial proteins are known to have the following common properties (20,21). 1) They all have a net basic charge and no acidic amino acids. 2) They share the common structural feature of being folded into an amphipathic α-helix with positively charged and hydrophobic faces which are thought to play a crucial role in transporting the proteins through the mitochondrial inner membrane electrophoretically in the membrane potential-dependent fashion (inside negative). The presequence of the mature subunit b has also these common structural features, as shown in Figure 3. Thus, the presequence could be the full

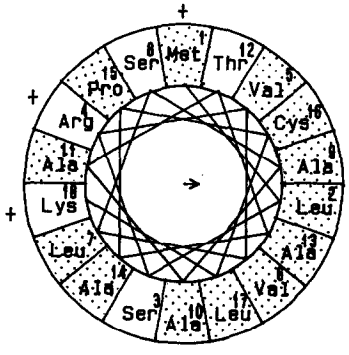

Fig. 3. Import signal peptide of subunit b plotted on helical wheels. Residues 1-18 of the presequence of subunit b were plotted onto helical wheel (24). The numbering starts at the proposed initiator methionine residue of the presequence of subunit b. Shaded residues indicate hydrophobic amino acids. (+), positively charged.

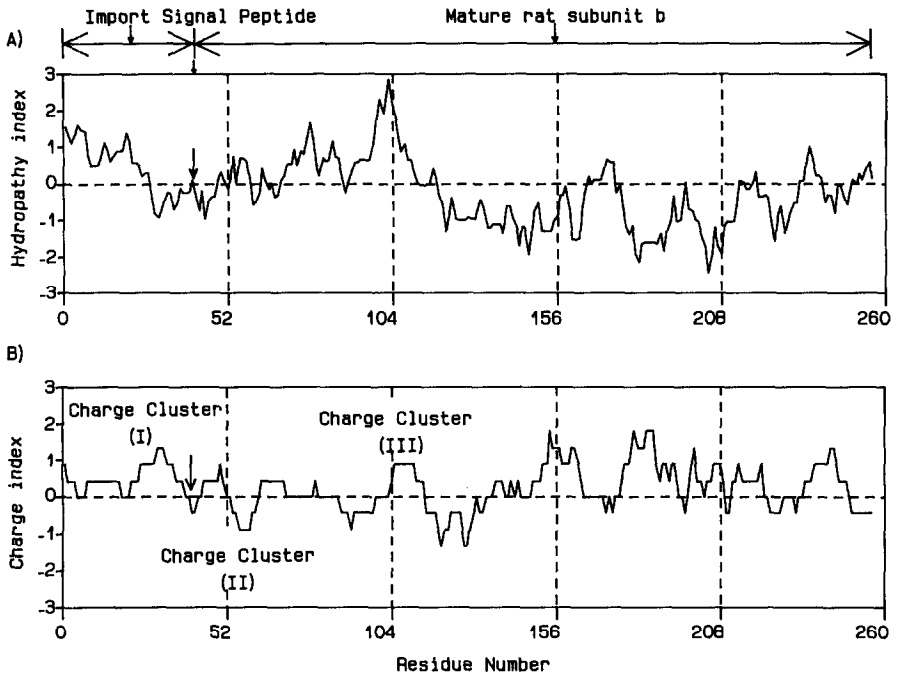

Fig. 4. Hydropathy and charge distribution plots of the amino acid sequence of the import precursor of subunit b of rat H^+-ATP synthase. The relative hydropathy of each residue was calculated by the method of Kyte and Doolittle (25) using a moving segment of 11 residues. The relative charge of each residue was calculated with a moving segment of 11 residues by using the following arbitrary charge index of amino acids: arginine, lysine, histidine, +5; glutamate, aspartate, -5; the other amino acids, 0. The relative charges of the N- and C-terminal amino acids were regarded as +5 and -5, respectively.

length of the import signal peptide of subunit b. This is the first report on the import signal peptide of mitochondrial subunit b.

The transmembrane domains of the premature subunit b were predicted from the hydropathy profile shown in Fig. 4(A) to extend from the premature residue 1 to 26 and residue 72 to 110 (the mature residue 30 to 68). Interestingly, there are three main unbalanced charge clusters in the N-terminus: positive charge cluster (I), negative charge cluster (II), and positive charge cluster (III), as shown in Fig. 4(B). These charge clusters could have a role to determine the orientation of the mature subunit b in F_o, in a membrane potential-dependent fashion. But, further studies are required to confirm this.

Finally, it should be noted that the mitochondrial subunit b does not conserve unbalanced negative charges in the C-terminus region that is known to be essential for ATP-dependent H^+-translocation by *Echerichia coli* H^+-ATP synthase and also for the binding of F_1 (22,23). This may indicate that the mitochondrial F_o-subunits differentiated functionally in the process of evolution. This could be a possible reason why the mitochondrial F_o is assembled from a much greater number of different polypeptides than the prokaryotic F_o-subunits.

References

1. Mitchell, P. (1961) *Nature* **191**, 423-427
2. Kagawa, Y. (1984) in *New Comprehensive Biochemistry* (Ernster, L., ed) Vol. 9, pp. 149-186, Elsevier Scientific Publishing Co., Amsterdam
3. Hatefi, Y. (1985) *Annu. Rev. Biochem.* **54**, 1015-1069
4. Futai, M., Noumi, T., & Maeda, M. (1989) *Annu. Rev. Biochem.* **58**, 111-36
5. Slater, E. C. (1987) *Eur. J. Biochem.* **166**, 489-504
6. Higuti, T. (1984) *Mol. Cell. Biochem.* **61**, 37-61 .
7. Oda, T., Futaki, S., Kitagawa, K., Yoshihara, Y., Tani, I., & Higuti, T. (1989) *Biochem. Biophys. Res. Commun.* **165**, 449-456
8. Muraguchi, M., Yoshihara, Y., Tunemitu, T., Tani, I., & Higuti, T. (1990) *Biochem. Biophys. Res. Commun.* in press.
9. Yoshihara, Y., Nagase, H., Yamane, T., Oka, H., Tani, I., & Higuti, T. (1989) *J. Biol. Chem.*, submitted
10. Maxam, A. M. and Gilbert, W. (1980) *Methods Enzymol.* **65**, 499-560.
11. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., & Rutter, W. J. (1979) *Biochemistry* **18**, 5294-5299.
12. Aviv, H., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1408-1412.
13. Fujiwara, T., Tanaka, K., Kumatori, A., Shin, S., Yoshimura, T., Ichihara, A., Tokunaga, F., Aruga, R., Iwanaga, S., Kakizuka, A., & Nakanishi, S. (1989) *Biochemistry* **28**, 7332-7340.
14. Hanahan, D., & Meselson, M (1980) *Gene* **10**, 63-67.
15. Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463-5467.
16. Walker, J. E., Runswick, M. J., & Poulter, L. (1987) *J. Mol. Biol.* **197**, 89-100.
17. Grantham, R., Gautier, C., Mercier, R., & Pave, A. (1980) *Nucleic Acids Res.* **8**, r49-r62.
18. Proudfoot, N. J., & Brownlee, G. G. (1978) *Nature* **263**, 211-214.
19. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 8857-8872.
20. Roise, D., Horvath, S. J., Tomich, J. M., Richards, J. H., & Schatz, G. (1986) *EMBO J.* **5**, 1327-1334.
21. von Heijne, G. (1986) *EMBO J.* **5**, 1335-1342.
22. Hoppe, J., Friedl, P., Schairer, H. U., Sebald, W., von Meyenburg, K., & Jørgenson, B. B. (1983) *EMBO J.* **2**, 105-110
23. Takeyama, M., Noumi, T., Maeda, M., & Futai, M. (1988) *J. Biol. Chem.* **263**, 16106-16112
24. Schiffer, M., & Edmundsen, A. B. (1967) *Biophys. J.* **7**, 121-135.
25. Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105-132.