Protein splicing of yeast VMA1-derived endonuclease via thiazolidine intermediates

Ryuta Mizutani,[a] Yasuhiro Anraku[b] and Yoshinori Satow[a*]

[a]Graduate School of Pharmaceutical Sciences, University of Tokyo, Hongo, Tokyo 113-0033, Japan, and [b]Department of Biosciences, Teikyo University of Science and Technology, Uenohara-machi, Yamanashi 409-0193, Japan. E-mail: satow@mol.f.u-tokyo.ac.jp

Protein splicing precisely excises out an internal intein segment from a protein precursor, and concomitantly ligates the N- and C-terminal extein polypeptides flanking the intein. A recombinant X10SNS bearing N- and C-extein polypeptides has been prepared for the intein endonuclease derived from the *Saccharomyces cerevisiae VMA1* gene. X10SNS has replacements of C284S, H362N, and C738S, and forms the intein and extein segments in the crystal lattice. The crystal structure of X10SNS revealed a linkage between the N- and C-extein segments, and showed that the C284 amino group of the resultant intein segment is in interaction with the G283 O atom of the N-extein segment. A mechanism for the final S→N acyl shift step proposes that a tetrahedral intermediate involves a five-memberred thiazolidine ring at G283-C738 junction. An oxyanion of the thiazolidine intermediate is to be stabilized by the C284 N atom.

Keywords: protein-splicing; *VMA1*-derived endonuclease; intein; extein; thiazolidine.

1. Introduction

Protein splicing is a post-translational process in which an intervening polypeptide (intein) is autocatalytically excised out from a precursor protein, and the flanking N- and C-terminal regions (N- and C-exteins) are ligated with a peptide bond. Evidence on protein splicing in nature came from the studies on the structure and expression of the *Saccharomyces cerevisiae VMA1* gene (Hirata *et al.*, 1990; Kane *et al.*, 1990). The gene contains an in-frame intervening intein sequence (the *VDE* open reading frame, Gimble & Thorner, 1992) and expresses a 120-kDa translational product of the Vma1 protein. The Vma1 protein auto-catalytically undergoes protein splicing to yield the 70-kDa Vma1 catalytic subunit of yeast vacuolar ATPase and the *VMA1*-derived site-specific homing endonuclease VDE that is also referred to as PI-*Sce*I endonuclease.

The protein-splicing reactions have been identified in a number of organisms from bacterial, archaeal, and eukaryotic species (Liu, 2000). The alignment of sequences related to the *VMA1* gene (Nogami *et al*,, 1997) shows that specific amino-acid residues at the N- and C-terminal junctions are highly conserved among inteins and exteins; the first residue of the intein is one of Cys and Ser, and the first residue of the C-extein is one of Cys, Ser, and Thr, which is placed immediately downstream of the highly conserved His-Asn sequence at the end of the intein. Key roles of these residues were pointed out by the site-directed mutagenesis studies and the biochemical analyses of the reaction intermediates on the truncated (Kawasaki *et al.*, 1996) and chimeric (Chong *et al.*, 1996) constructs of VDE.

The protein splicing of the Vma1 protein *in vitro* is folding dependent and requires only six N-extein and four C-extein residues (Kawasaki *et al.*, 1996). Information on conformations of these N- and C-extein residues in the Vma1 protein, therefore, is essential for understanding of the protein splicing reaction. A series of the recombinant VDE proteins with segments of 11 N-extein and 10 C-extein residues from the *VMA1* sequence has been reported (Mizutani *et al*., 2002). In these recombinants, mutations decreasing the reaction speed were introduced; C284S, H362N and C738S in X10SNS; C284S, H362N, N737S and C738S in X10SSS; and H362N and C738S in X10CNS (numbering of the amino-acid sequence refers to that for the Vma1 protein, Hirata *et al*., 1990).

The X10CNS recombinant undergoes the reaction and yields splicing products. The recombinant X10SNS whose crystal structure is reported here is a spliceable precursor which undergoes the splicing reaction in the crystal lattice. Based on the X10SSS crystal obtained as a stable precursor, the reaction mechanism involving a thiazolidine intermediate has been proposed for the initial N→S acyl shift step of the reaction (Mizutani *et al*., 2002). The stepwise scheme proposed for protein splicing is summarized in figure 1. Through the initial step of the reaction, a thioester intermediate is formed at the conserved C284. This X10SNS study is aimed at the elucidation for the structural basis of the final S→N acyl shift step.

2. Materials and methods

2.1. Crystallization and diffraction data collection

The X10SNS recombinant was expressed and purified as described previously (Mizutani *et al*., 2002). X10SNS was

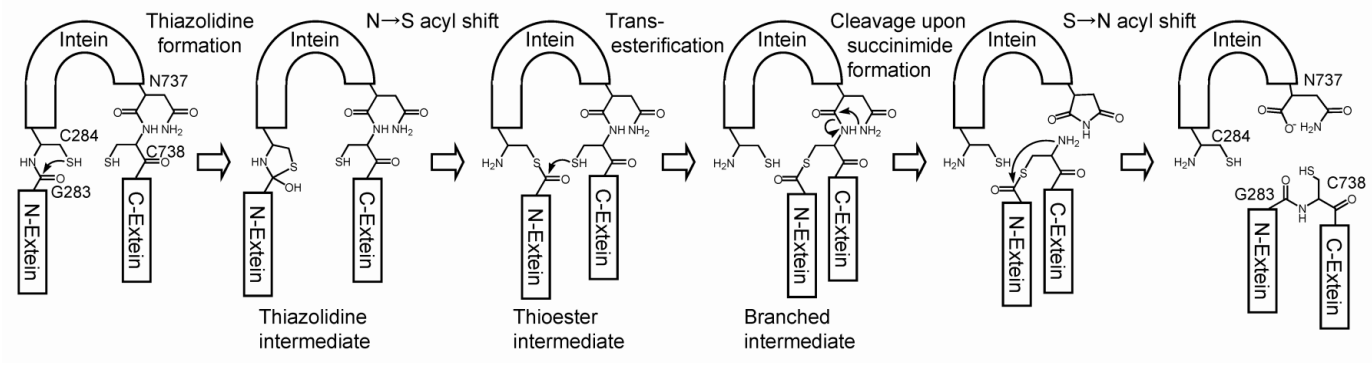


Figure 1 Schematic diagram for the mechanism of the protein-splicing reaction. The thiazolidine intermediate at the G283-C284 junction is formed and resolved into the thioester intermediate. The nucleophilic attack on the thioester by the conserved C738 side-chain gives the branched intermediate. In the next step, the intervening region from C284 to N737 is excised out by the peptide bond cleavage upon succinimide formation at the conserved N737. The transient product then undergoes the final S→N acyl shift from a thioester to an amide bond. This step is further examined as shown in figure 3.

crystallized with the hanging-drop vapor-diffusion method using a 16% (w/v) polyethyleneglycol (PEG) 6000 reservoir solution containing 0.1 M BisTris-HCl pH 6.2, 10 mM 2-mercaptoethanol, 10 mM $MgCl_2$, and 1 mM $CdCl_2$. Protein droplets were prepared by mixing equal volumes (5 µl) of the protein and reservoir solutions. Plate-shaped crystals grew to sizes of 0.12 mm x 0.07 mm x 0.03 mm in 26.5 hours at 20°C. These crystals were dialyzed against a cryoprotectant solution containing 8% (w/v) PEG 400 for 2 hours and 25% for additional 2 hours, and then flash-cooled in a 110-K stream of nitrogen gas. The crystals diffract synchrotron X-rays to 2.9 Å resolution, and belong to space group *P*1 with lattice constants of a = 66.4 Å, b = 70.4 Å, c = 58.0 Å, α = 100.4°, β = 98.7°, and γ = 78.9°. An asymmetric unit contains two X10SNS molecules, and hence the value for the crystal volume per unit molecular mass is obtained as 2.4 $Å^3$/Da which corresponds to a solvent content of 0.49 (v/v).

Diffraction intensities from the X10SNS crystals were collected by use of synchrotron X-rays at the BL40B2 station of SPring-8, Japan. With the X-ray wavelength of 0.7000 Å and the crystal-to-detector distance of 500 mm, 135 diffraction patterns were recorded on the R-AXIS IV^{++} detector (Rigaku) with an oscillation range of 2.0° and an exposure time of 100 sec. Reflections of 50,211 in total were reduced into 20,700 unique reflections with R_{merge} of 0.051 using programs DENZO and SCALEPACK (Otwinowsky & Minor, 1997). The data set is 83% complete in the 30 - 2.9 Å shell, and 60% complete in the 3.0-2.9 Å shell. Statistics for the data collection are summarized in table 1.

2.2. Structure determination and refinement

The structural model of the X10SSS VDE (Mizutani *et al.*, 2002) without extein residues was subjected to the rigid body refinement with 8-3.5 Å data using the program CNS (Brünger *et al.*, 1998). At the end of this refinement, the crystallographic *R*-factor dropped from 0.44 to 0.36. Further refinements were carried out with the simulated annealing procedure using CNS. After each refinement cycle, the model was fitted to electron-density maps with the program TURBO-FRODO (Roussel & Cambillau, 1991), and then refined with the simulated annealing. The N- and C-extein residues, which were removed from the initial VDE model, were built into the model. The loop regions at residues 338-350, 553-562, and 651-659 located away from the splicing site were observed in low electron-densities protruding into the solvent region, and were not included in the final model that has been deposited with Protein Data Bank.

Table 1

Diffraction data collection and structure refinement statistics.

Resolution range (Å)	30.0-2.9
Number of observed reflections	50,211
Number of unique reflections	20,700 (1,519[a])
Completeness	0.830 (0.601[a])
R_{merge}[b]	0.051 (0.141[a])
Number of protein molecules per asymmetric unit	2
Number of protein atoms	6,710
Number of water molecules	161
R (R_{free}[c])	0.214 (0.283)
Rms deviations from idealities	
Bond lengths (Å) and angles (°)	0.012 / 1.72
Dihedral and improper angles (°)	24.3 / 1.01

[a] Numbers in parentheses are for the resolution range of 3.0-2.9 Å.
[b] $R_{merge} = \Sigma_h \Sigma_j^{Nh} | I_j(h) - \langle I \rangle(h) | / \Sigma_h \Sigma_j^{Nh} I_j(h)$, where intensity $I_j(h)$ is measured N_h times for a reflection h.
[c] R_{free} is calculated for the 5% subset of the unique reflections.

3. Results and discussion

3.1. Structure of the X10SNS VDE

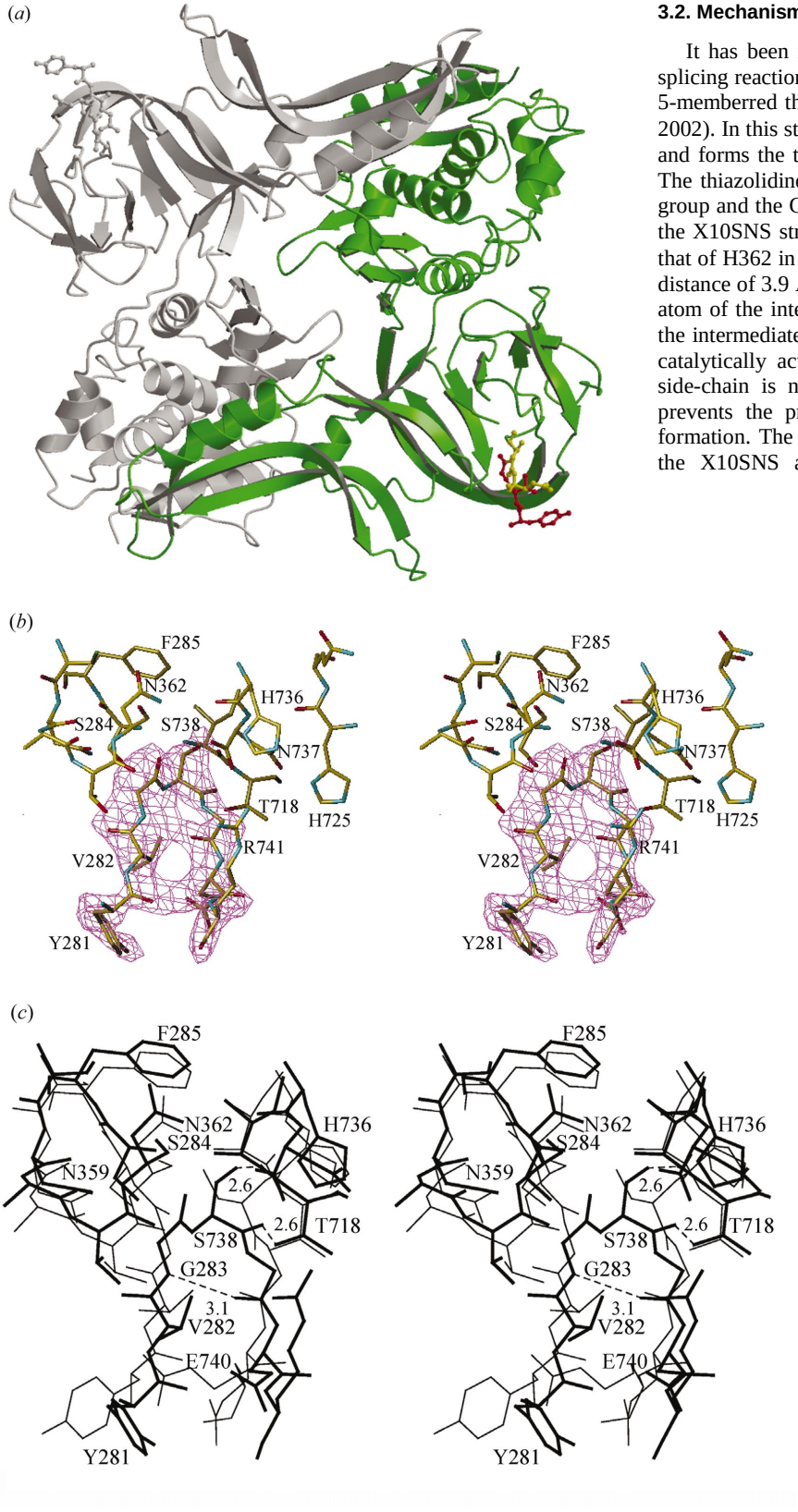
The crystal structure of X10SNS was refined to the *R*–factor of 0.214 at 2.9 Å resolution. The statistics on the structure refinement are shown in table 1. The structures of X10SNS molecules A and B in the asymmetric unit are virtually identical. An rms positional difference between these molecules is 0.46 Å for the splicing region of the domain II (residues 284-368, 438-462, and 700-737). The splicing site is situated at the bottom of the β-sheet region of the domain II (figure 2a). Electron densities for the N- and C-extein residues are continuously observed for residues Y281-G283, and S738-R741 in each molecule (figure 2b). The results of electrophoretic and mass spectrometric analyses (Mizutani *et al.*, 2002) indicated that the N- and C-extein segments of X10SNS are excised from the precursor in the crystal lattice, and that the excised segments from the X10CNS recombinant are ligated auto-catalytically. As for the ligated extein segments whose structures are fitted in the electron densities, splicing junctions (residues 282-285, 359-362, and 736-739) give a small rms positional difference of 0.45 Å between molecules A and B. It has been reported that N737 of the excised intein turns up in an approximately equimolar amount of asparagine and succinimide forms (Chong *et al.*, 1996). Electron densities for the residue 737 are interpretable either as an asparagine or as a succinimide, but an asparagine structure is well fitted into the densities.

In the splicing site, many atomic contacts are observed among hydrophobic residues V282, P324, and Y714. The region of residues G283, S738, and G739 is in an atypical turn conformation: with dihedral angles of ϕ = -130° and ψ = 36° for G283, ϕ = 107° and ψ = -9° for S738, and ϕ = -150° and ψ = 163° for G739. This turn structure involves a hydrogen bond with a distance of 3.1 Å between the G283 N and G739 O atoms for molecule B. The main-chain conformation of S738 is stabilized by a hydrogen bond between its O and T718 Oγ1 atoms (2.6 Å). Residues 739 - 741 are in the β-strand conformation, forming a short antiparallel β-sheet along with the N-extein residues V282 and G283, with a hydrogen bond between the G283 N and G739 O (3.1 Å) and an interaction between the Y281 O and R741 N (3.9 Å). The formation of this β-sheet brings the residues of the N- and C-terminal junctions in close proximity; making the G283 and S738 conformation suitable for undergoing the S→N acyl shift. The side chains of H736 and N737 are in the vicinity of the 283-738 linkage; a distance between the G283 O and N737 O atoms is 3.6 Å, and that between the H736 Nδ1 and N737 O atoms is 2.7 Å.

Between the X10SNS and X10SSS structures, the splicing region of the intein domain II gives a small rms difference of 0.42 Å. This indicates that the overall structure of the splicing region is virtually unalterable by the splicing reaction, although the extein itself gives an rms difference of 0.84 Å between the X10SNS and X10SSS. Major differences are only observed in the N-extein residues Y281, V282, and G283, which are shifted by 1.2 Å toward the outside of the VDE molecule (figure 2c) in the X10SNS structure.

Splicing-defective mutants of the GST-Vma1 fusion protein, with a replacement of either G283V, C284Y, H362L, V403D, S639P, N737S, or N737K, have been identified from a random mutagenesis study (Kawasaki *et al.*, 1997). Any replacement in the catalytic residues of C284, H362, and N737 brings an immediate effect on the splicing activity. The G283V replacement would cause a steric hindrance with the catalytic H362 side-chain. The hydrophobic side-chain of V403 is in interaction with that of I442 which is located in an anti-parallel β-sheet formed by Y437-E443

and E363-P369 strands. The V403D replacement would also affect the spatial arrangement of H362 that is located upstream of the E363-P369 strand in the β-sheet. The S639P replacement abolishes a hydrogen bond between the S639 N and S635 O atoms and disrupts the helix formation at H626-L640, resulting in misfolding of the domain II.



3.2. Mechanism of the splicing reaction

It has been reported that the initial N→S acyl shift step of the splicing reaction proceeds via a tetrahedral intermediate containing a 5-memberred thiazolidine ring as shown in figure 1(Mizutani *et al.*, 2002). In this step, the C284 Sγ thiolate attacks on the G283 C atom, and forms the thiazolidine intermediate at the G283-C284 junction. The thiazolidine intermediate is then resolved into the C284 amino group and the G283-C284 thioester linkage. The N362 side-chain in the X10SNS structure is in the conformation virtually equivalent to that of H362 in the generic VDE: the N362 Nδ2 atom is located at a distance of 3.9 Å from the G283 O atom, and 3.4 Å from the oxygen atom of the intermediate modelled in the X10SNS structure. Hence the intermediate is possibly stabilized by the N362 side-chain in the catalytically active H362N mutants. However, the replaced N362 side-chain is not capable of protonating the C284 N atom, and prevents the protonation that leads to the thioester intermediate formation. The decreased activities for the splicing reaction in both the X10SNS and X10SSS recombinants are ascribable to the replaced N362 side-chain.

In the subsequent transesterification step, the neutral C284 amino-group formed by the opening of the thiazolidine ring activates the C738 Sγ atom, which then attacks on the G283 thioester carbonyl carbon. The resultant tetrahedral intermediate is resolved into the C284 thiolate and a branched intermediate in which the N-extein acyl group is linked to the C738 side-chain. This branched intermediate is then cleaved off from the intein region upon the succinimide formation at N737. In the X10SNS structure, the location of the H725 side-chain is suitable for the succinimide formation, with a distance of 3.1 Å between the H725 Nδ1 and N737 Nδ2 atoms.

Figure 2 Structure of the X10SNS VDE protein. (a) X10SNS VDE molecules in the asymmetric unit. Molecules A and B are viewed along the non-crystallographic twofold axis. The N- and C-extein residues are illustrated as stick models. Molecule B is highlightened by colors; red, N-extein residues; light green, intein residues; yellow, C-extein residues. This figure is produced with program MOLSCRIPT (Kraulis, 1991). (b) Stereo view of the electron-density map superposed on the extein residues of molecule B. For the calculation of this difference-Fourier map, extein residues 281-283 and 738-741 were omitted from the model, and then the resultant model was further refined. The map is contoured at 3 σ level using TURBO-FRODO (Roussel & Cambillau, 1991). (c) Stereo drawings of the splicing site of X10SNS and X10SSS. X10SNS residues are drawn in bold lines, and X10SSS residues in thin lines with TURBO-FRODO. Hydrogen bonds involving the N- and C-extein residues are indicated by broken lines, with their donor-acceptor distances in Å.

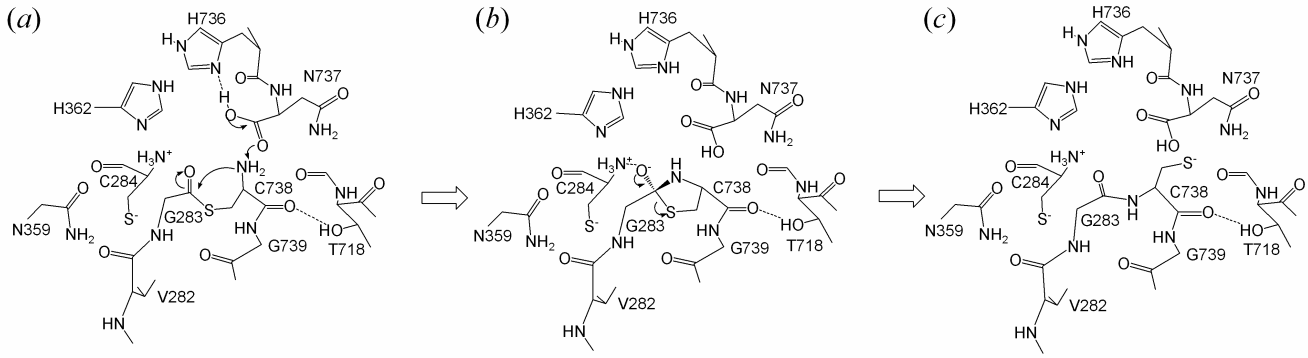


Figure 3 Mechanism for the final S→N acyl shift step of the protein-splicing reaction via a tetrahedral intermediate with a thiazolidine structure. N737 is shown as a carboxyl form which is protonated by H736. Broken lines represent possible hydrogen bonds. (a) The nucleophilic attack on the G283 carbonyl C' by the C738 neutral amino-group initiates the S→N acyl shift. (b) A tetrahedral intermediate containing a thiazolidine ring is formed. An oxyanion at G283 is stabilized by the C284 N atom. (c) The thiazolidine intermediate is resolved into the C738 thiolate and a peptide bond between G283 and C738.

In the final S→N acyl-shift step of the protein splicing, the thioester structure is further transformed into a peptide-bond linkage between G283 and C738. Figure 3 shows this step proposed on the basis of the splicing site of the X10SNS structure. As the thioester linkage is chemically unstable, the S→N acyl shift reaction would proceed more rapidly in the hydrophobic environment of the splicing site, and the spontaneous hydrolysis be prevented. In the beginning of the S→N acyl shift, the neutral C738 amino-group donates its proton to the carbonyl oxygen of N737, and thus formed C738 N nucleophile attacks on the G283 C' atom. The resultant tetrahedral intermediate has a thiazolidine ring structure which is also formed in the initial N→S acyl shift step. The C284 amino group of X10SNS is located at a distance of 2.9 Å from the G283 O of the N-extein segment. This distance indicates that an oxyanion of this second thiazolidine intermediate is stabilized by the C284 N atom. The thiazolidine intermediate is then resolved into a thiolate and the peptide bond between the N- and C-exteins. Since the X10SSS, X10SNS, and X10CNS recombinants supposedly retains common precursor and intermediate structures, the transesterification and succinimide formation steps in the middle of the splicing reaction still remain to be clarified.

This work was supported by the RFTF program 96L00505 from Japan Society for the Promotion of Science to YS, by the Scientific Research Program B033005 on Priority Areas from the Japanese Ministry of Education, Science, Sports, and Culture to YS, and by a Grant-in-Aid for Advanced Scientific Research on Bioscience/Biotechnology Areas from the MEXT of Japan to YA. Diffraction data was collected at SPring-8 (Proposal 2000A0356 and 2000A0486 by RM).

References

Brünger, A.T., Adams, P.D., Clore, G.M., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. & Warren, G.L. (1998). *Acta Cryst. Sect. D* **54**, 905-921.
Chong, S., Shao, Y., Paulus, H., Benner, J., Perler, J.B. & Xu, M.Q. (1996). *J. Biol. Chem.* **271**, 22159-22168.
Gimble, F.S. & Thorner, J. (1992). *Nature* **357**, 301-306.
Hirata, R., Ohsumi, Y., Nakano, A., Kawasaki, H., Suzuki, K. & Anraku, Y. (1990). *J. Biol. Chem.* **265**, 6726-6733.
Kane, P.M., Yamashiro, C.T., Wolczyk, D.F., Neff, N., Goebl, M. & Stevens, T.H. (1990). *Science* **250**, 651-657.
Kawasaki, M., Makino, S., Matsuzawa, H., Satow, Y., Ohya, Y. & Anraku, Y. (1996). *Biochem. Biophys. Res. Commun.* **222**, 827-832.
Kawasaki, M., Nogami, S., Satow, Y., Ohya, Y. & Anraku, Y. (1997). *J. Biol. Chem.* **272**, 15668-15674.
Kraulis, P.J. (1991). *J. Appl. Cryst.* **24**, 946-950.
Liu, X.Q. (2000). *Annu. Rev. Genet.* **34**, 61-76.
Mizutani, R., Nogami, S., Kawasaki, M., Ohya, Y., Anraku, Y. & Satow, Y. (2002). *J. Mol. Biol.* **316**, 919-929.
Nogami, S., Satow, Y., Ohya, Y. & Anraku, Y. (1997). *Genetics* **147**, 73-85.
Otwinowsky, Z. & Minor, W. (1997). *Meth. Enzymol.* **276**, 307-326.
Roussel, A. & Cambillau, C. (1991). TURBO-FRODO Version 1.1 release B, Faculte de Medecine Nord, Marceille, France.