Biochemical and Biophysical Research Communications **286,** 171–175 (2001)
doi:10.1006/bbrc.2001.5354, available online at http://www.idealibrary.com on IDEAL®

Assessment of Substrate Specificity of Hepatitis G Virus NS3 Protease by a Genetic Method

Yong Jae Lee, Hara Kang, Seong-Hwan Rho, Soo Hyun Eom, and Woo Jin Park[1]

Department of Life Science and National Research Laboratory of Proteolysis, Kwangju Institute of Science and Technology, Kwangju 500-712, Korea

Received July 10, 2001

The RNA genome of hepatitis G virus (HGV) encodes a large polyprotein that is processed to mature proteins by viral-encoded proteases. The HGV NS3 protease is responsible for the cleavage of the HGV polyprotein at four different locations. No conserved sequence motif has been identified for the cleavage sites of the NS3 protease. To determine the substrate specificity of the NS3 protease, amino acid sequences cleaved by the NS3 protease were obtained from randomized sequence libraries by using a screening method referred to as GASP (Genetic Assay for Site-specific Proteolysis). Based on statistical analyses of the obtained cleavable sequences, a consensus substrate sequence was deduced: Gln-Glu-Thr-Leu-Val ↓ Ser, with the scissile bond located between Val and Ser. The relevance of this peptide as a cleavable substrate was further supported by molecular modeling of the NS3 protease. Our result would provide an insight on the molecular activity of the NS3 protease and may be useful for the design of substrate-based inhibitors.

Hepatitis G virus (HGV or GBV-C) is a transfusion-transmissible agent identified in the sera of hepatitis patents (1, 2). The HGV genome is a 9.4-kb positive-stranded RNA molecule which contains an open reading frame (ORF) that encodes a polyprotein of 2900 amino acids (2). Sequence comparison of the viral polyproteins revealed that HGV is closely related to HCV and two non-human viruses GBV-A and GBV-C, and thus classified as a member of the family *Flaviridae* (1–3).

The putative nonstructural region of the HGV polyprotein resembles the NS2–NS3–NS4A–NS4B–NS5A–NS5B structure of the HCV polyprotein, and is processed to mature proteins by the NS2 and NS3 proteases (2, 4, 5). In particular, HGV NS3, a chymotrypsin-like serine protease, has been shown to cleave at NS3/NS4A, NS4A/NS4B, NS4B/NS5A, and NS5A/NS5B junctions (2, 4). Thus, the mode of HGV polyprotein cleavage in the nonstructural region appears to be very similar to that of HCV. However, the substrate specificity of the HCV and HGV NS3 proteases is quite different.

We have previously developed a genetic method (GASP; Genetic Assay for Site-specific Proteolysis) that was shown to be an efficient tool for the study of substrate specificity of proteases (6, 7). A similar method was used to detect caspase activity in yeast (8, 9). In this study, we investigated the substrate specificity of the HGV NS3 protease by using GASP. Our result should help us to understand the molecular mechanism of NS3 protease activity and open the way for the design of substrate-based inhibitors.

MATERIALS AND METHODS

Construction of pADH-SteSubLex and pGAL-HGVNS3. The construction of two backbone vectors, pADH-SteLex and pGAL, was described previously (7). Oligonucleotides (5′-CAT GGC AAG AAA CCT TGG CCT CCT TCT CTT ACA TTT GGC-3′ and 5′-GAT CGC CAA ATG TAA GAG AAG GAG GCC AAG GTT TCT TGC-3′) encoding the 5A/5B junction were synthesized, annealed, and inserted into the *Nco*I–*Bam*HI sites of pADH-SteLex, resulting in pADH-SteSubLex. For randomization of residues in the substrate region, degenerate oligonucleotides were synthesized as described previously (6). Library complexity with randomization in one residue (P2, P1, P1′) was about 1000 each and that with two residues randomization (P3–P2, P5–P4) was about 100,000 each. A DNA fragment encoding the whole NS3/4A region was amplified from HGV (kindly provided by Dr. J. P. Kim) and inserted into the *Eco*RI–*Sal*I sites of pGAL, resulting in pGAL-HGVNS3

Yeast transformation and selection for GASP. The *Escherichia coli* XL-1 Blue strain (*supE144 hsdR17 recA1 endA1 gyrA96 thi-1 lac*$^-$ *F′[proAB*$^+$ *lacI*^{q} *lacZi*Δ*M15 Tn10(tet*^{r}*)]*) (10) was used for the manipulation of plasmid DNA. The yeast strain EGY48 (*MAT*α, *ura3, his3, trp1, LexA*$_{OP(X6)}$*-LEU2*) (11) was used for study of the substrate specificity of the HGV NS3 protease by GASP. The pADH-SteSubLex libraries with randomized substrate sequences were introduced by the lithium acetate method (12) into EGY48 harboring

[1] To whom correspondence should be addressed at Department of Life Science, Kwangju Institute of Science and Technology, 1 Oryong-dong, Puk-gu, Kwangju, 500-712, Korea. Fax: +82-62-970-2484. E-mail: wjpark@kjist.ac.kr.

0006-291X/01 $35.00

the protease vector, pGal-HGVNS3. The yeast cells were plated onto non-selective plates lacking Trp, His, Ura, and incubated for 3 days at 30°C. The cells were then replica plated onto GAL-inducible selective plates containing 2% galactose and 1% raffinose but lacking Trp, His, Ura, and Leu, and onto GAL-repressible selective plates containing 2% glucose instead of galactose and raffinose. Colonies growing only on the GAL-inducible selective plates were picked and subjected to further characterization such as X-gal staining (13) and nucleotide sequencing.

DNA sequencing. Direct colony PCR was performed to obtain the substrate regions of the selected positive colonies. The primer set was (forward) 5′-CTCGT TCCCT TTCTT CCTTG TTTC-3′ and (backward) 5′-TACCT GTCTG GCTGA TGTG-3′. The amplified DNA fragments were sequenced with the backward primer by using an ABI Prism Model 377 automated DNA sequencer.

Modeling of the substrate specificity of the NS3 protease. Sequence alignment with AlignMaster revealed that 9 PDB HCV NS3 protease sequences displayed 37.15% sequence identity with the HGV NS3 protease. Among these, the cofactor and inhibitor complex form, 1DXWA was selected as a template for structural modeling of the NS3 protease. The first model was obtained by submitting the HGV NS3 protease sequence (amino acids 904–1108) to the SWISS MODEL server (14) with the template structure 1DXWA. The binding of peptide substrate and NS4A cofactor was manually built by manipulating side chains using O (15). Then, the built-in energy minimization function of the Swiss-PDB viewer was applied repeatedly using GROMOS96B force field and a few inappropriate side chains were fixed using the built-in loop database. The color figure was prepared by using the Swiss PDB viewer and the POV-RAY tracing program.

RESULTS AND DISCUSSION

GASP to detect the activity of the HGV NS3 protease. In this study, we sought to determine the substrate specificity of the NS3 protease by utilizing a recently developed screening method referred to as GASP (Genetic Assay for Site-specific Proteolysis) (6, 7). The principle of GASP is shown in Fig. 1A. A fusion protein, SteSubLex, was generated in which a transcription factor, LexA-b42, is linked to the truncated cytoplasmic domain of an integral membrane protein, STE2, by a short linker containing a substrate sequence of the NS3 protease. The rationale was that: (i) in the absence of the NS3 protease activity, LexA-b42 remains anchored to the cytoplasmic membrane (left panel); (ii) in the presence of the NS3 protease, the substrate sequence is cleaved, resulting in the release of LexA-b42 and its translocation into the nucleus where it activates reporter genes (right panel).

The construction of two backbone vectors, pADH-SteLex and pGAL, was described previously (7). We further constructed a substrate vector, pADH-SteSubLex, in which an oligonucleotide encoding a peptide spanning the P8–P6′ region of the NS5A/5B cleavage site of the polyprotein (E A R Q E T L A ↓ S F S T I W) was inserted between the STE2 and LexA-b42 coding sequences of pADH-SteLex (Fig. 1B). The expression of the fusion proteins, SteLex and SteSubLex, was driven by the constitutive ADH promoter. A DNA fragment encoding the NS3 protease was amplified by PCR and inserted into pGAL, resulting in the

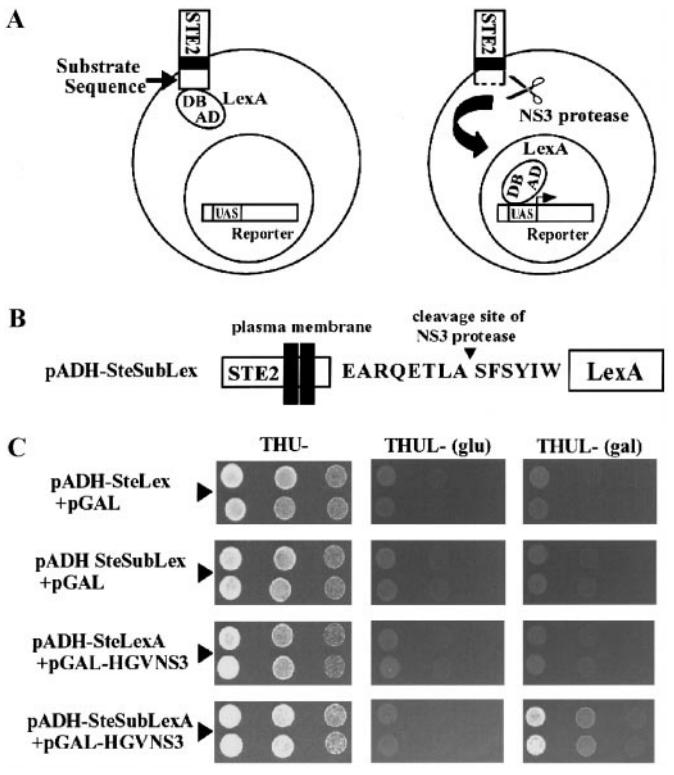

FIG. 1. Utilization of GASP (Genetic Assay for Site-specific Proteolysis) to detect the activity of the NS3 protease. (A) A fusion protein, SteSubLex, is expressed in a yeast strain that contains LexA-b42-inducible reporter genes. LexA-b42 is anchored in the plasma membrane via an integral membrane protein (STE2) and a substrate sequence for the NS3 protease (left). When the NS3 protease is introduced, the substrate sequence is cleaved, resulting in the release of LexA-b42 and subsequent activation of reporter genes (right). (B) The structure of SteSubLex is shown. The substrate sequence (EARQETLASFSYIW) is derived from the 5A/5B cleavage site of the HGV polyprotein. (C) Yeast cells transformed with substrate and protease vectors, as indicated, were diluted 10-fold each and spotted onto nonselective and selective plates. Only the transformants harboring pADH-SteSubLex and pGAL-HGVNS3 (row 4) showed strong X-gal staining and were able to grow on the selective plates containing galactose but not glucose. This indicates the specific cleavage of the substrate sequence by the NS3 protease in yeast.

protease vector pGAL-HGVNS3. The expression of NS3 protease was under the control of the inducible GAL promoter. A *Leu2* mutant reporter strain EGY48 was transformed with both pADH-SteSubLex and pGAL-HGVNS3, and spotted onto Leu-free plates containing either glucose (nonselective) or galactose (selective) (Fig. 1C). The transformants showed strong X-gal staining (data not shown) and excellent growth on the selective plates containing galactose, but not at all on the plates containing glucose (row 4). This indicated that the expression of the NS3 protease was strictly regulated by the GAL promoter and that the substrate sequence in SteSubLex was specifically cleaved by the NS3 protease. Further, results from control experiments performed with the blank back-

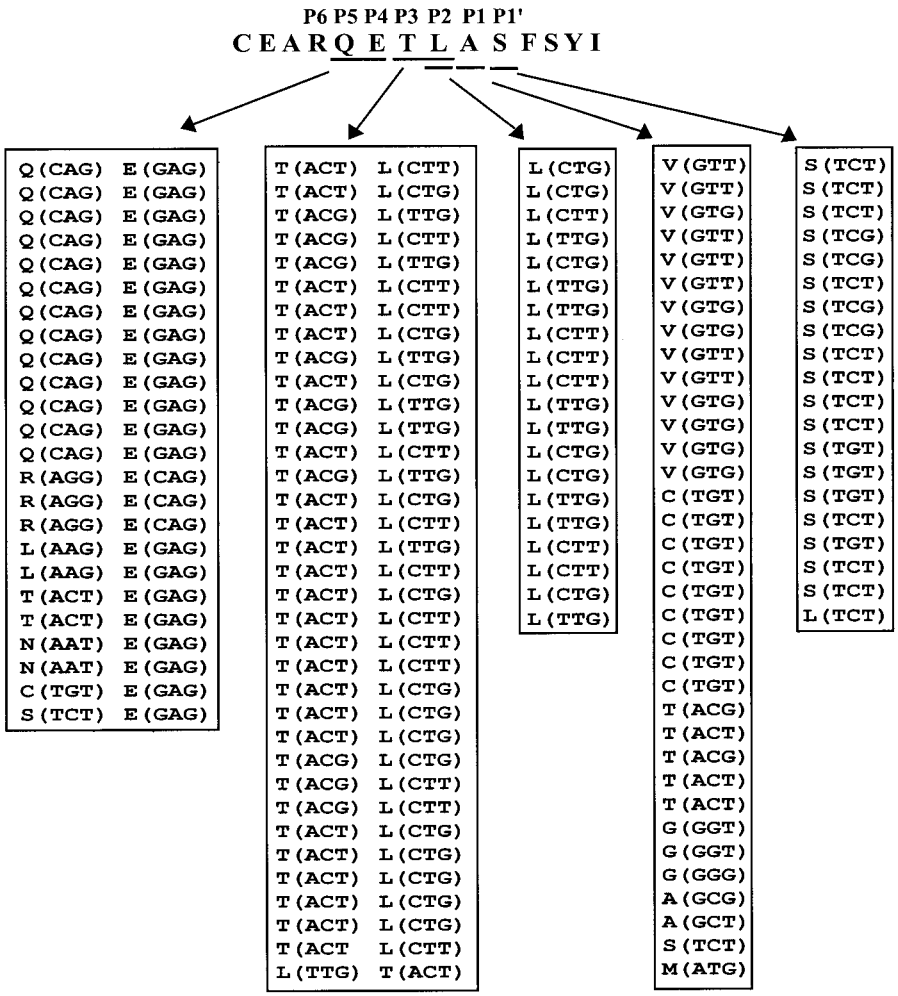

FIG. 2. The underlined sequences, QE, TL, L, A, and S were randomized in separate experiments. The obtained cleavable amino acid sequences with their corresponding nucleotide sequences are shown in each box.

bone vectors, pADH-SteLex and pGAL, indicated that proteolysis was highly specific (rows 1, 2, and 3).

Screening for substrate sequences cleaved by the NS3 protease. Good substrate sequences for the NS3 protease were selected basically as described previously (6, 7). Briefly, we randomized the P5–P4, P3–P2, P2, P1, and P1′ positions of the NS5A/5B junction in separate experiments. Each randomized sequence was expected to contain any one of the 20 natural amino acids. The yeast cells harboring pGAL-HGVNS3 were transformed with these randomized substrate libraries and plated onto plates lacking Trp and His (*Trp1* and *His3* are the maker genes for the substrate and protease vectors, respectively), and then replica plated onto selective plates lacking Trp, His, and Leu (*Leu2* is the reporter gene that is activated by LexA-b42) and containing either glucose or galactose. They were incubated for 3–4 days at 30°C. Colonies growing only on the selective plates containing galactose but not glucose were supposed to harbor amino acid sequences that were specifically cleaved by the NS3 protease, and thus these colonies were selected. The substrate regions of the selected positive colonies were then amplified by PCR, and directly subjected to nucleotide sequencing. The cleavable sequences are shown in Fig. 2. It was striking to see that locations P4, P3, P2, P1′ were highly conservative and that the occupied residues were those of the native NS5A/5B region. Our assay system might be less sensitive setting relatively high threshold for the detection of NS3 protease activity, so that suboptimal residues were eliminated during the selection. In this regard, it is notable that another native substrate sequence, the NS4B/5A junction, was not cleaved in our system (data not shown). Gln was a predominant residue in P5. Val was the most frequently occurring residue at P1 and its frequency surpassed that of native Ala. Taken together, our results suggest that Gln-Glu-Thr-Leu-Val ↓ Ser with the scissile bond located between Val and Ser is the most favored cleavage sequence for the NS3 protease.

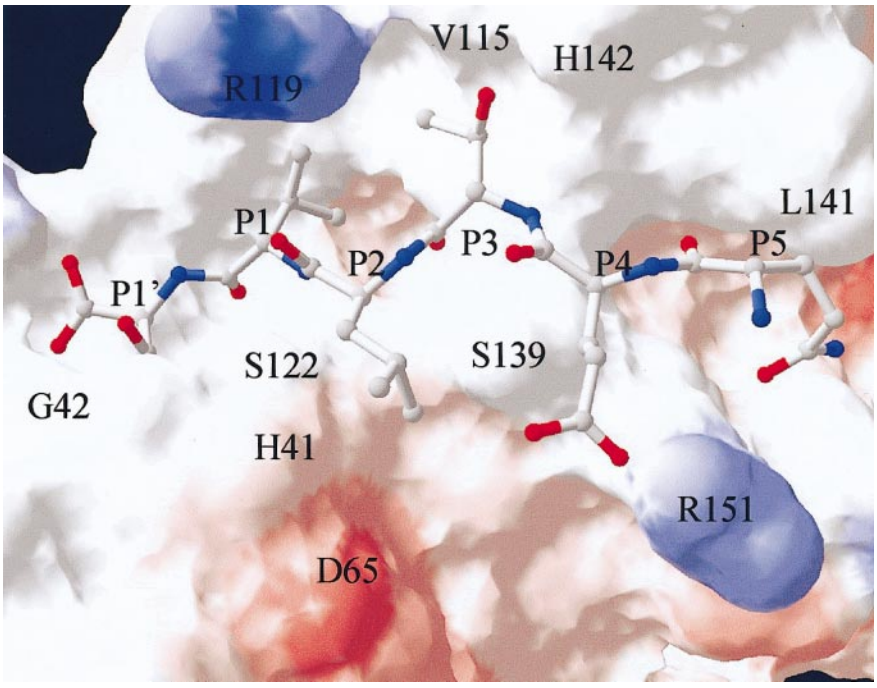

FIG. 3. Comparative modeling of the consensus substrate sequence QETLVS and the NS3 protease. The active site of the NS3 protease domain is shown as a surface representation and the substrate peptide QETLVS is shown as a ball-and-stick model with carbon atoms in gray, oxygen in red, and nitrogen in blue. Electrostatic surface potentials are contoured from −6 (red) to 6 (blue) kT/e.

This is exactly the same sequence found at the NS5A/5B junction except that Val occurs at P1 instead of Ala, implying that the NS5A/5B site has the closest possible match to the optimal cleavage sequence.

Modeling of substrate binding to the NS3 protease. To see if the consensus substrate sequence obtained from this study fits reasonably within the active site of the NS3 protease, we performed modeling of the substrate binding to the NS3 protease using the structure of the HCV NS3 protease as a template (see Materials and Methods) (Fig. 3). First, the catalytic triad His41, Asp65, and Ser122 and the oxyanion hole formed by amide groups from Gly120, Ser121 and positively charged Arg119 are well matched within the structural context of the serine protease family (16). The small groove at S1′ defined by Gly42 requires a small polar residue at P1′ and forces the substrate backbone to make an abrupt turn after a possible hydrogen bond with the backbone amide of Gly42. Hydrophobic residues are generally acceptable at P1. In the HCV NS3 protease, Phe154 determines the specificity of the P1 site in which Cys is the most favorable residue (17). In HGV, however, Ser27 takes the place of this bulky side chain, thus resulting in a preference for hydrophobic residues such as Val and Ala at P1. Interestingly, Leu is a predominant residue at P2, although it does not appear to be required spatially. The side chain of Leu at P2 is positioned above the His41 imidazole moiety. It is known that the hydrogen bond between catalytic Asp-His residues becomes unstable when it is exposed to solvent and conversely extremely stable when sheltered from solvent ($\Delta G'' > 10$ kcal/mol). Therefore, Leu at P2 may be critical for maintaining the integrity of the catalytic triad as observed with other serine proteases including the HCV NS3 protease (18). The backbone amide and carbonyl groups of Thr at P3 and Val140 are involved in intermolecular, anti-parallel β-like hydrogen bonding. The side chain of P3 does not seem to be in a defined conformation, rather it is free to move in the space delimited by His142, Arg119, and Val115. This is not unusual because P3 residues are frequently solvent-exposed or do not interact with the enzyme in a number of serine protease-inhibitor complexes. P4 and P5 are also highly conserved. One possible explanation is that Glu at P4 and Gln at P5 may interact with Arg151 via a salt bridge and a hydrogen bond, respectively. In conclusion, the interpretations drawn from the modeling of substrate binding to the NS3 protease are largely consistent with the results of our GASP analysis of the NS3 protease.

CONCLUDING REMARKS

The method used in this study, GASP, has been successfully used for detecting caspase activity (8, 9), and determining the substrate specificity of HCV (hepatitis C virus) NS3 protease (6). In this study, we have successfully determined the substrate specificity of

HGV NS3 protease by using GASP. The result implies that GASP might be used for the study of other viral proteases. GASP offers a more reliable and simpler method than phage display or standard *in vitro* biochemical methods in the study of substrate specificity of proteases. The consensus substrate sequence obtained from this study might be useful for the design of substrate-based NS3 protease inhibitors.

ACKNOWLEDGMENTS

This work was supported by grants from the National Research Laboratory, Molecular Medicine Research Group Program (98-J03-02-03-A-01) from the Ministry of Science and Technology, and BK21 program from the Ministry of Education, Korea. The authors are grateful to Dr. Jungsuh P. Kim (Genelabs Technologies, Inc.) for kindly providing the HGV NS3 protease gene.

REFERENCES

1. Simons, J. N., Leary, T. P., Dawson, G. J., Pilot-Matias, T. J., Muerhoff, A. S., Schlauder, G. G., Desai, S. M., and Mushahwar, I. K. (1995) *Nat. Med.* **1,** 564–569.
2. Linnen, J., Wages, J., Jr., Zhang-Keck, Z. Y., Fry, K. E., Krawczynski, K. Z., Alter, H., Koonin, E., Gallagher, M., Alter, M., Hadziyannis, S., Karayiannis, P., Fung, K., Nakatsuji, Y., Shih, J. W., Young, L., Piatak, M., Jr., Hoover, C., Fernandez, J., Chen, S., Zou, J. C., Morris, T., Hyams, K. C., Ismay, S., Lifson, J. D., Kim, J. P., *et al.* (1996) *Science* **271,** 505–508.
3. Simmonds, P., and Smith, D. B. (1999) *J. Virol.* **73,** 5787–5794.
4. Belyaev, A. S., Chong, S., Novikov, A., Kongpachith, A., Masiarz, F. R., Lim, M., and Kim, J. P. (1998) *J. Virol.* **72,** 868–872.
5. Charrel, R. N., Attoui, H., De Micco, P., and de Lamballerie, X. (1999) *Biochem. Biophys. Res. Commun.* **255,** 432–437.
6. Kim, S. Y., Park, K. W., Lee, Y. J., Back, S. H., Goo, J. H., Park, O. K., Jang, S. K., and Park, W. J. (2000) *Anal. Biochem.* **284,** 42–48.
7. Kang, H., Kim, S. Y., and Park, W. J. (2001) *Mol. Cells* **11,** 263–266.
8. Hawkins, C. J., Wang, S. L., and Hay, B. A. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 2885–2890.
9. Steiner, H., Pesold, B., and Haass, C. (1999) *FEBS Lett.* **463,** 245–249
10. Bullock, W. O., Fernandez, J. M., and Short, J. M. (1987) *BioTechniques* 5, 376–379
11. Estojak, J., Brent, R., and Golemis, E. A. (1995) *Mol. Cell. Biol.* **15,** 5820–5829.
12. Gietz, D., St Jean, A., Woods, R. A., and Schiestl, R. H. (1992) *Nucleic Acids Res.* **20,** 1425.
13. Rose, M., and Botstein, D. (1983) *Methods Enzymol.* **101,** 167–180.
14. Guex, N., and Peitsch, M. C (1997) *Electrophoresis* **18,** 2714–2723
15. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A* **47,** 110–119
16. Love, R. A., Parge, H. E., Wickersham, J. A., Hostomsky, Z., Habuka, N., Moomaw, E. W., Adachi, T., and Hostomska, Z. (1996) *Cell* **87,** 331–342.
17. Kim, J. L., Morgenstern, K. A., Lin, C., Fox, T., Dwyer, M. D., Landro, J. A., Chambers, S. P., Markland, W., Lepre, C. A., O'Malley, E. T., Harbeson, S. L., Rice, C. M., Murcko, M. A., Caron, P. R., and Thomson, J. A. (1996) *Cell* **87,** 343–355.
18. Barbato, G., Cicero, D. O., Cordier, F., Narjes, F., Gerlach, B., Sambucini, S., Grzesiek, S., Matassa, V. G., De Francesco, R., and Bazzo, R. (2000) *EMBO J.* **19,** 1195–1206.