Engineering and Chemical Synthesis of a Transmembrane Protein: The HCV Protease Cofactor Protein NS4A

Elisabetta Bianchi,[§] Raffaele Ingenito,[§] Reyna J. Simon,[||,†] and Antonello Pessi*[,§]

Istituto di Ricerche di Biologia Molecolare P. Angeletti (IRBM), Via Pontina Km 30.600 00040 Pomezia (Rome), Italy
Gryphon Sciences, 250 East Grand Avenue South San Francisco, California 94080

Received January 19, 1999

Although integral membrane proteins are encoded by about a quarter of the genes in both prokaryotic and eukaryotic organisms, based on analysis of open reading frames[1] very little is known about the structural basis of their function. Failure of recDNA-based expression systems to produce the high amounts of purified protein required for biochemical and structural studies is generally due to the properties of the protein itself, which almost invariably tend to form, when expressed at high levels, intracellular precipitates in the form of inclusion bodies;[2] these in turn are difficult to refold to active proteins.[3] Even when successfully prepared, membrane proteins do not lend themselves easily to physicochemical and structural studies due to their low solubility.

We had to address both problems, i.e., synthetic difficulties and poor solubility, in the course of our studies on the obligatory cofactor protein, NS4A, of the serine protease NS3 of Hepatitis C virus, and wish to report on our progress. NS3 is an HCV-encoded protein that includes a serine protease and a helicase. The serine protease domain (henceforth NS3pro), which is contained in the N-terminal 180 aa, is required for maturation of the viral polyprotein and therefore is the target of an intense drug discovery effort worldwide; to cleave several junctions of the HCV polyprotein, however, NS3pro must form a complex with another virally encoded protein, the cofactor NS4A.[4] More recently, in fact, the HCV protease is being referred to as the NS3/4A protease.[5] NS4A is a 54-aa protein that has been predicted to be a type I transmembrane (TM) protein[6] and is known to target NS3 to the endoplasmic reticulum.[7] The TM helical domain is located in the 20 N-terminal residues,[6] while the central region (residues 21−34) in the form of a synthetic peptide (Pep4A) is the minimal domain required for NS3 activation. Complex formation between the protease domain and Pep4A has been characterized kinetically and structurally, including X-ray crystal structure.[8−9] Several aspects of the mechanism of activation of NS3pro by NS4A, however, remain unsolved, in particular the role of the N- and C-termini of the cofactor. For example, the observed in vitro stability of the NS3/Pep4A complex is considerably lower than what has been observed in in vitro translation experiments for the complex between NS3 and full-length NS4A.[8]

So far any attempt at preparing NS4A by recDNA failed,[10] and to date no protein is available for study. We have attempted the total chemical synthesis of the protein by the Fmoc/t-Bu method. NS4A is the prototype of a "difficult sequence" according to Kent's definition.[11] Its sequence is shown in Figure 1, together with the Kyte−Doolittle hydropathy plot[12] and the calculated $\langle p \rangle$ value for predicted synthetic problems.[13] Not surprisingly, even with use of extended coupling times and HATU as activator,[14] the synthesis produced a highly heterogeneous crude product; moreover, the protein proved so insoluble, even in highly solvating media like DMSO or in the presence of high concentrations of chaotropic agents such as 6 M guanidine hydrochloride, that it prevented an effective chromatographic purification.

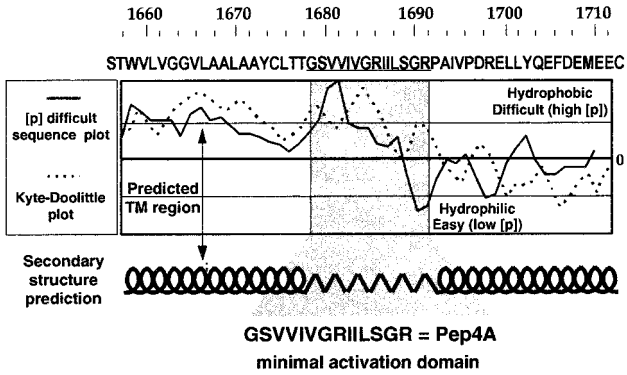

Figure 1. Sequence,[4] hydrophobicity plot,[13] "difficult-sequence" plot,[14] and predicted secondary structure[27,28] of the HCV NS4A protein. The numbering above the sequence refers to the HCV polyprotein. The minimal sequence required for NS3 activation is underlined in the sequence and highlighted in gray. The predicted transmembrane region[6] (amino acids 1−20) is also shown.

We therefore decided to engineer NS4A into a more soluble protein by addition of a 3-lysine tail, a strategy that we had successfully pursued in the past.[15] Derivatization at the N-terminus improved solubility in solvent systems suitable for chromatographic purification, like 20−50% AcOH, but was insufficient to confer the desired solubility in media suitable for physicochemical studies (solubility in 10 mM Tris buffer, pH 7, is <1 μM). We then pursued tagging at both termini and, since the C-terminus of NS4A is part of a site for intra as well as intermolecular cleavage by NS3,[4] we decided to elongate NS4A past the cleavage site and to introduce a 4-lysine tail in position P7′ of the NS4A/4B cleavage site. This protein finally showed the desired properties. It is monomeric[16] and forms a fully active complex upon incubation with the NS3 protease domain. Solubility in 10 mM Tris buffer, pH 7, is >100 μM.

The small amount of NS4A and the engineered analogues necessary for the above work were produced by stepwise chemical synthesis, which yielded the needed, minute amounts of purified

* Corresponding author: Istituto di Ricerche di Biologia Molecolare P. Angeletti (IRBM). Phone: +39-06-91093445. Fax: +39-06-91093654. E-mail: pessi@irbm.it.
[§] Istituto di Ricerche di Biologia Molecolare P. Angeletti (IRBM).
[||] Gryphon Sciences.
[†] Present address: MetaXen, 280 East Grand Avenue, South San Francisco, CA 94080.

(1) (a) Wallin, E.; von Heijne, G. *Protein Sci.* **1998**, *7*, 1029−1038. (b) Saraste, M.; Walker, J. E. *Curr. Opin. Struct. Biol.* **1998**, *8*, 477−479.
(2) Mitraki, A.; King, J. *Biotechnology* **1989**, *7*, 690−697.
(3) Jaenicke, R.; Rudolph, R. **I**n *Protein Structure: A Practical Approach*; Creighton, T. E., Ed.; IRL Press: Oxford, 1989; pp 191−223.
(4) Failla, C.; Tomei, L.; De Francesco, R. *J. Virol* **1994**, *68*, 3753−3760.
(5) Kwong, A. D. *Curr. Opin. Infectious Dis.* **1997**, *10*, 485−490.
(6) Rost, B.; Casadio, R.; Fariselli, P.; Sander, C. *Protein Sci.* **1995**, *4*, 521−533.
(7) Hijikata, M.; Mizushima, H.; Tanji, Y.; Komoda, Y.; Hirowatari, Y.; Akagi, T.; Kato, N.; Kimura, K.; Shimotohno, K. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10773−10777.
(8) Bianchi, E.; Urbani, A.; Biasiol, G.; Brunetti, M.; Pessi, A.; De Francesco, R.; Steinkühler, C. *Biochemistry* **1997**, *36*, 7890−7897.
(9) (a) 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.; Thomson, J. A. *Cell* **1996**, *87*, 343−355. (b) Yan, Y.; Li, Y.; Munshi, S.; Sardana, V.; Cole, J.; Sardana, M.; Steinkühler, C.; Tomei, L.; De Francesco, R.; Kuo, L.; Chen, Z. *Protein Sci.* **1998**, *7*, 837−847.
(10) R. De Francesco, personal communication.
(11) Kent, S. B. H. *Annu. Rev. Biochem.* **1988**, 57, 957−989.
(12) Kyte, J.; Doolittle, R. F. *J. Mol. Biol.* **1982**, *157*, 105−132.
(13) Milton, R. C. de L.; Milton, S. C. F.; Adams, P. A. *J. Am. Chem. Soc.* **1990**, *112*, 6039−6045. The software used in our calculations is contained in the package "Peptide Companion", written by M. Lebl, G. Lebl, and V. Krchnak, commercially available from Peptides Int.: Louisville, KY.
(14) Carpino, L. A.; El-Faham, A.; Minor, C. A.; Albericio, F. *J. Chem. Soc., Chem. Commun.* **1994**, 201−203.
(15) Bianchi, E.; Venturini, S.; Pessi, A.; Tramontano, A.; Sollazzo, M. *J. Mol. Biol.* **1994**, *236*, 649−659.

10.1021/ja9901617 CCC: $18.00
Published on Web 08/07/1999

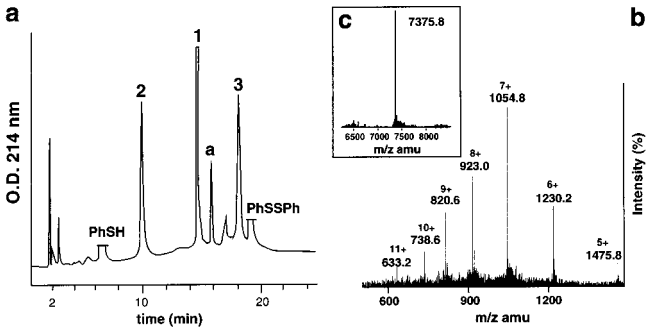

Figure 2. Native chemical ligation of engineered NS4A protein. (a) Analytical HPLC of the reaction after 29 h. Conditions: Column, Vydac C-4, 150 × 4.6 mm, 5 μm, 300 Å; eluent A, 0.1% (acq) TFA; eluent B, 0.1% TFA in MeCN, isocratic 30% B (5′) then linear gradient 30−60% B (20′). Sample preparation: see Supporting Information. Peaks 1 and 2 correspond to precursor fragment peptides **1** and **2** (see text), peak a is activated peptide **1**−COSPhenyl formed by reaction of peptide **1** with thiophenol. (b) Ion-spray mass spectrometry of HPLC-purified peak 3 acquired on an API-365 (PE-SCIEX) triple-quadrupole mass spectrometer; the multiple charge states arise from protonation of a single species of MW 7376. (c) Hypermass reconstruction of the spectrum shown in part b [calculated (average isotopic composition) 7376.83 daltons, found 7375.8 daltons].

product. Once the desired form of NS4A had been established, however, we looked for a better method to produce larger amounts of homogeneous protein for full physicochemical and biological studies. Synthetic problems during stepwise assembly were mostly due to failure to efficiently incorporate the 16 N-terminal residues corresponding to the putative TM helix of NS4A (Figure 1). Native chemical ligation,[17] a method in which two unprotected peptides, a C-terminal thioester and an N-terminal cysteine-containing fragment, are joined together by a chemoselective reaction in aqueous solution, seemed ideally suited for our purpose, since the required cysteine residue is conveniently located close to one end of the N-terminal TM domain. Moreover, ligation chemistry is compatible with the detergents and chaotropes which we anticipated as necessary to solubilize the reaction components. Native chemical ligation of engineered NS4A was thus carried out between a 19-aa N-terminal fragment corresponding to the putative transmembrane helix plus the three-lysine tail (peptide **1**, KKK-STWVLVGGVLAALAAY-COSR) and a 47-aa fragment corresponding to the rest of the molecule (peptide **2**, CLTTGSVVIV-GRIILSGRPAVIPDREVLYQEFDEMEECASHLPYKKKK-NH_2).

Peptide **1** was assembled by using stepwise Boc/Benzyl chemistry with optimized protocols[18] on a peptide thioester resin analogous to the method used by Aimoto and co-workers,[19] while peptide **2** was prepared by Fmoc/t-Bu chemistry.[20] The most important parameter for success of the ligation reaction was the choice of the reaction medium. On one side, high concentrations of reactants were recommended to avoid sluggish reaction kinetics, possibly worsened by the presence of a less-preferred (with respect to Ala or Gly) tyrosine residue as the C-terminal thioester;[21] on the other side, their solubility and that of the target product, lysine tails notwithstanding, was rather low. We therefore ran the reaction in a buffer (25 mM HEPES, pH 7) containing a detergent frequently used for crystallization of membrane proteins, i.e., β-octyl-glucoside.[22] The ligation was started by addition of 1% thiophenol to a solution of peptides **1** and **2** (2.0 and 2.07 mM, respectively) to promote conversion of the C-terminal α-thioester to the more reactive phenyl α-thioester.[23] The reaction was very slow, being only about 50% complete in 90 h (Figure 2). We decided to stop the ligation at this point by acidification with TFA, and proceeded directly to HPLC purification of the target protein. On the basis of starting material, the overall yield of the ligation was 40%. Both peptides **1** and **2** were recovered in the same purification step, and could be successfully recycled in a new reaction. The protein produced by chemical ligation was as active as the one previously obtained by stepwise assembly.

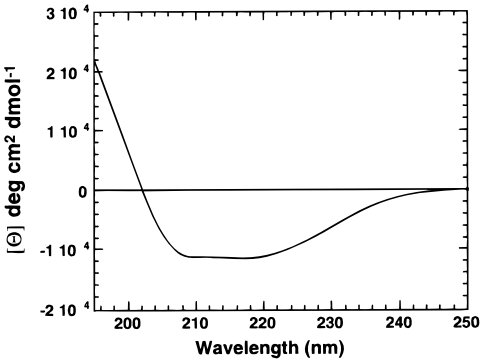

Figure 3. Baseline-subtracted far-UV circular dichroism spectrum of synthetic engineered NS4A protein. Buffer, 10 mM HEPES, 15% glycerol, 1% β-octyl-glucoside, 6 mM urea; protein concentration (by quantitative amino acid analysis), 25 μM; cell length, 0.05 cm; temperature, 20 °C.

Preliminary structural analysis of the engineered NS4A protein was performed by circular dichroism in the same buffer conditions used for biological activity studies (Figure 3). Reassuringly, a good agreement was observed between the secondary structure predicted[24,25] and calculated[26] from the spectral data: predicted, 42% α-helix, 21% β-sheet, 37% remainder; calculated from spectrum, 36% α-helix, 22% β-sheet, 42% remainder. More studies are in progress on the biochemical and structural characterization of the complex between NS3 and engineered, full-length NS4A and these will be reported elsewhere.

In conclusion, we have been able to engineer a type I TM protein into a soluble form while preserving the secondary structure and the biological activity. Moreover, a high-yield synthesis of the protein, which proved elusive by stepwise SPPS, was achieved by native chemical ligation by using an appropriately selected detergent in the reaction medium. We envisage a good potential for further application of this combined strategy to the study of transmembrane proteins.

Acknowledgment. We thank N. Dimasi, S. Di Marco, M. Cerretani, S. Altamura, and C. Steinkühler for help with the characterization of the protein, F. Bonelli and S. Orru' for mass spectrometry, R. Cortese for continuous support, and M. Sollazzo and V. G. Matassa for critical reading of the manuscript.

Supporting Information Available: Detailed procedures for the stepwise synthesis and purification of a representative full-length NS4A protein and peptides **1** and **2**, for the chemical ligation step, and for the purification of the final product (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA9901617

(16) Assessed by dynamic light scattering on a Dyna Pro-801 instrument, using a 1 mM solution of NS4A in 10% glycerol, 2% β-octyl-glucoside.
(17) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776−779.
(18) Schnolzer, M.; Alewood, P.; Alewood, D.; Kent, S. B. H. *Int. J. Pept. Protein Res*. **1992**, *40*, 180−186.
(19) (a) Hojo, H.; Aimoto, S. *Bull. Chem. Soc. Jpn*. **1991**, *64*, 111−117. (b) Hojo, H.; Kwon, Y.; Kakuta, Y.; Tsuda, S.; Tanaka, I.; Hikichi, K.; Aimoto, S. *Bull. Chem. Soc. Jpn*. **1993**, *66*, 2700−2706. (c) Hojo, H.; Aimoto, S. *Bull. Chem. Soc. Jpn*. **1993**, *66*, 3004−3008.
(20) Atherton, E.; Sheppard, R. C. *Solid-phase peptide synthesis, a practical approach*; IRL Press: Oxford, 1989.
(21) Kent, S. B. H.; Muir, T. W.; Dawson, P. E. WO 96/34878, **1996**, p 14.
(22) Buffer composition: 25 mM HEPES, pH 7.0, 2% (v/v) β-octyl-glucoside, plus 1% (v/v) thiophenol; buffer combinations which proved unsuccessful due to target protein precipitation were 2−6 M guanidine hydrochloride with or without addition of organic cosolvents such as MeCN, dioxane, and DMF.
(23) Dawson, P. E.; Churchill, M. J.; Reza Ghadiri, M.; Kent, S. B. H. *J. Am. Chem. Soc*. **1997**, *119*, 4325−4329.
(24) Rost, B.; Sander, C. *J. Mol. Biol*. **1993**, *232*, 584−599.
(25) Rost, B.; Sander, C.; Schneider, R. *CABIOS* **1994**, *10*, 53−60.
(26) Andrade, M. A.; Chacon, P.; Merelo, J. J.; Moran, F. *Protein Eng*. **1993**, *6*, 383−390.