0273-2289/96/6101-02—0085$08.00

Protein Engineering of *De Novo* Protein with Predesigned Structure and Activity

DMITRY A. DOLGIKH,*[,1] ANDREI E. GABRIELIAN,[1] VLADIMIR N. UVERSKY,[2] AND MICHAEL P. KIRPICHNIKOV[1]

[1]*Institute of Molecular Biology, Russian Academy of Sciences, 117984, Moscow, Russia; and* [2]*Institute of Protein Research, Russian Academy of Sciences, Moscow Region, Russia*

ABSTRACT

The *de novo* protein albebetin has been engineered (*J. Mol. Biol.* 1992, **225**, 927–931) to form a predesigned tertiary fold that has not yet been observed in natural proteins. Analysis of albebetin expressed in a cell-free system and in *Escherichia coli* revealed its compactness, relative stability, and the secondary structure close to the predesigned one. The blast-transforming biological activity of human interferon was grafted to albebetin by attachment of an eight amino acid interferon fragment to the N-terminus of albebetin next to its first methionine residue. The chimeric protein was expressed in a wheat germ cell-free translation system and tested for its structural properties, receptor binding, and biological activity. According to the tests, albebetin incorporating the active interferon fragment has a compact and relatively stable structure, and binds the murine thymocyte receptor effectively. It activates the blast transformation reaction of thymocyte cells even more efficiently than human interferon at low concentrations.

Index Entries: *De novo* proteins; protein design; protein structure; protein expression.

INTRODUCTION

Design of *de novo* proteins, which ensures the desirable three-dimensional structures and functions, is one of the most intriguing challenges of modern protein engineering, the best way to test our understanding of

*Author to whom all correspondence and reprint requests should be addressed. E-mail: dolgikh@imb.ac.ru.

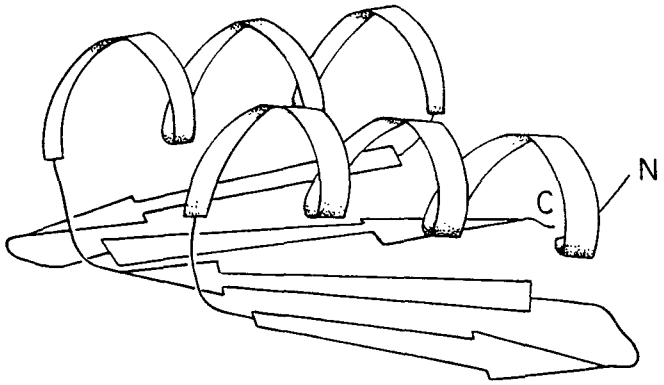

Fig. 1. Predesigned three-dimensional structure of *de novo* protein albebetin.

the principles of protein structure and folding. The first successful engineering of a *de novo* protein, namely four-helix bundle, had been reported seven years ago (*1*), and then a number of attempts were made in this direction during the past few years (for a recent review *see 2*). Most of the design structures mimicked the structures of some natural proteins. The next attractive step in this field can be construction of proteins having not only the predesigned structure, but possessing a biological activity.

In 1992 (*3*) a *de novo* protein albebetin was engineered with a predesigned tertiary fold (Fig. 1) that does not contradict any structural rule, but has never been observed in natural proteins. We have expressed albebetin in a cell-free translational system in nanogram amounts. Its initial testing showed that the *de novo* protein is relatively stable and nearly as compact as natural proteins (*3*).

In this article, we present our results on albebetin expression in *Escherichia coli* and its investigation and introduction of an eight amino acid biologically active human interferon fragment into albebetin in order to obtain *de novo* protein with a grafted biological activity.

MATERIALS AND METHODS

Genetic Engineering

Standard methods (*4*) were used for genetic engineering procedures, such as DNA digestion and analysis with restriction endonucleases, ligation, and transformation of *E. coli*, preparation and purification of DNA fragments, and so forth. Restriction endonucleases were purchased from Promega and Fermentas (Lithuania), and T4 DNA ligase was from Pharmacia. The oligonucleotides were synthesized on a Gene Assembler, Pharmacia, and purified by FPLC according to the Pharmacia manual.

Protein Expression

mRNA for albebetin expression in the wheat germ cell-free translational system was transcripted from a plasmid by phage SP6 RNA-polymerase as described (*5*). The protein was expressed in the cell-free system (*6*). For albebetin expression in *E. coli*, we used a fusion expression system from New England Biolabs (NEB) and cloned the gene into plasmid pMALc after the N-terminal part of the maltose binding protein gene. Expression, purification, and cleavage of the fusion protein with factor Xa protease was performed according to the NEB manual, as well as the final affinity protein purification using the FPLC column chromatography on amylose binding resin.

Size-Exclusion Chromatography

Size-exclusion chromatography was done using the FPLC equipment and a Superose 12 column (Pharmacia) calibrated by protein standards. The flow rate was 20 mL/h, and 0.2-mL fractions were collected. The radioactivity incorporated into the polypeptide was measured after precipitation with trichloracetic acid and plotted vs the fraction number. The Stokes radii values were determined as described (*7,8*).

Circular Dichroism Measurements

Circular dichroism spectrum was measured in 20 m*M* Tris-HCl buffer, pH 7.8, and 10 m*M* NaCl using a Jasco-600 spectropolarimeter. The protein concentration was 0.4 mg/mL.

BInding to Receptors and Biological Testing

For analysis of the inhibitory effects, murine thymocytes (10^7–10^8 cells/mL) were incubated for 40 min at 4°C with the unlabeled human interferon or albebetin incorporating the interferon fragment and 4.4×10^{-11} *M* of the radiolabeled interferon octapeptide in 50 m*M* Tris-HCl buffer containing bacitracin (0.1 mg/mL), EGTA (1 m*M*), phenylmethylsulfonyl fluoride (0.6 mg/mL), and the soybean trypsin inhibitor (0.1 mg/mL). The incubation was stopped by rapid filtration through GF/B glass filters (Whatman), and the filters were rinsed, dried, and their radioactivity was counted with a Mini Gamma Counter (LKB). The results were plotted as a percent of the specific binding vs log of the competitior concentration, and the IC_{50} values corresponding to the 50% inhibition were determined graphically (*see* Fig. 6 later in article). The inhibition constant (K_i) was calculated as described elsewhere (*9*). The blast-transforming activity was assayed as described (*10*).

RESULTS AND DISCUSSION

Albebetin Design

The principles of albebetin design have been described in detail elsewhere (*11*). Therefore, only the main features are summarized here. α-Helices and β-regions were designed on the basis of the physical theory of protein secondary structure (*12–15*) and checked by the computer program ALB (*14*). Valine and leucine residues were inserted into proper positions in α- and β-regions, correspondingly, in order to ensure the formation of a nonpolar protein core. α-Helices and β-strands were designed to have the same length of 1.8 nm and α-and β-regions were spearated by α- and β-breaking glycines and prolines. These connections had minimal lengths in order to minimize the number of possible structures. The designed amino acid sequence was carefully checked, taking into account the polar character of the protein surface in order to prevent the protein from aggregation. The density of packing in albebetin globule was analyzed using the Corey, Pauling, and Koltun (CPK) space-filling model of albebetin and then by computer graphics using the program "Insight II" (Biosym).

Albebetin Expression and Investigation

The design of albebetin gene and its expression in a cell-free translational system have been described (*3*). Initial testing of the protein by gel-filtration, urea-gradient electrophoresis, and limited proteolysis showed that it is compact and relatively stable, and possesses a tertiary structure close to or at least not contradicting the predesigned one (*3*). For albebetin expression in *E. coli*, we used the fusion system with maltose-binding protein as described in Materials and Methods. The expressed protein has been tested by the SDS polyacrylamide gel and urea-gradient electrophoresis, and size-exclusion chromatography (data not shown), and revealed the same properties as albebetin obtained in a cell-free translational system.

For analysis of a secondary structure of the *de novo* protein, we used the circular dichroism (CD) spectroscopy in the far-UV region. The CD spectrum of albebetin shown in Fig. 2 reflects a high content of the secondary structure and is similar to the spectra of natural proteins containing α- and β-structures. The quantitative analysis of the spectrum (*16*) showed that albebetin contains 28% α- and 40% β-structure. These values are close to the predesigned ones (30 and 36%, respectively), indicating that the secondary structure design was successful, and therefore one can suppose that the tertiary structure of albebetin also corresponds to the predesigned one.

Thus, albebetin is compact, structured, and possesses the secondary structure similar to the predesigned one. These data are not sufficient to conclude whether this *de novo* protein has a rigid tertiary structure typical

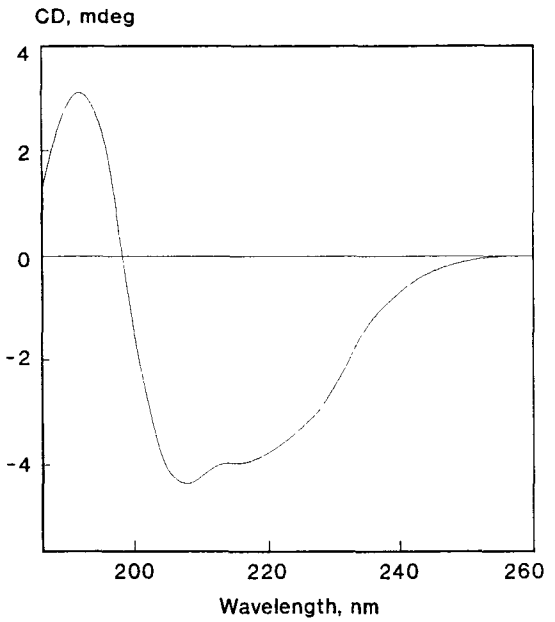

Fig. 2. Circular dichroism spectrum of albebetin in 20 m*M* Tris-HCl buffer, pH 7.8, 10 m*M* NaCl.

of native natural proteins or it is in the molten globule state like most of *de novo* proteins designed up to now. A more detailed study, including the NMR measurements or X-ray analysis, can clarify the question (the NMR study of albebetin is in progress in the laboratory of C. Dobson, Oxford University, UK). Nevertheless, the data obtained allowed us to modify albebetin based on its predesigned structure, namely, to introduce a biological activity into the *de novo* protein.

A *DE NOVO* PROTEIN WITH A GRAFTED BIOLOGICAL FUNCTION

Principles of Grafting

To transfer a biological activity to a *de novo* protein, one can try to imitate a natural protein active center. However, this is rather difficult, especially in the case of a spatially arranged active center (like enzyme catalytic sites), because the exact three-dimensional structure of a *de novo* protein is not known, and, as a consequence, it is impossible to reproduce the original spatial structure of such a center. On the other hand, the contiguous functionally important site possessing biological activity *per se* as a peptide may simply be inserted into a *de novo* protein. It is only necessary to check whether the overall structure of chimeric protein is maintained and whether the introduced sequence is able to form a conformation required for the biological functioning.

There are two obvious ways to insert an active site into a protein, i.e., replacement and attachment. According to the first approach, one could find a protein sequence, which is most homologous (or may be analogous, i.e., similar in arrangement of hydrophobic/hydrophilic amino acids) to the active site of interest, and to insert it by replacing the inherent protein sequence. The closer the homology, the greater in principle, is the probability of successful design. Nevertheless, it was shown that even the same sequences could acquire various conformations in different proteins (*17*). The second approach is to insert the active sequence into an exposed part of the molecule. No protein sequence is substituted, and the local conformational changes in the exposed site would not seriously disturb the protein structure. The inserted sequence would have enough flexibility to acquire the active conformation. The loop regions and the termini of the protein molecule are suitable for this kind of a design.

Biological Activity to Be Grafted to Albebetin

In search for a biological activity that could be transferred to albebetin, the following objectives have been taken into account:

1. An easy test for the detection of this biological activity;
2. Since albebetin is an acidic protein, a basic functional fragment would be preferable; and
3. The active fragment should be relatively short not to disturb the albebetin structure.

We compared the amino acid sequence of albebetin with the data base (*18*) for more than 150 functionally important fragments of proteins described in the literature, but no satisfactory homology with known functional sites was found. As a result, a construction of biologically active albebetin is only possible by attachment of an active fragment to the protein-exposed site. The octapeptide LKEKKYSP (Leu-Lys-Glu-Lys-Lys-Tyr-Ser-Pro) was chosen as a candidate. It corresponds to the human interferon α_2 sequence 131–138, binds strongly to murine thymocyte receptors, and activates the thymocyte blast transformation at a concentration as low as $10^{-11}M$ (*10*). It is known that the fragment 131–138 is located in a loop of the interferon molecule and has a considerably reduced affinity to the receptor compared to the peptide itself (*10*). On the other hand, it has been shown that the receptor binding properties of this peptide disappear after linking amino acids to its N-terminus and reduce after linking the Cys residue to the C-terminus (*10*). Taking this into account, we have chosen attachment of this octapeptide to albebetin N-terminus as the way of its incorporation into the protein molecule. The octapeptide LKEKKYSP corresponds to all the requirements listed above, its binding to receptors is easy to test owing to the high-affinity and the sensitive assay, it is basic, and is relatively short and highly hydrophilic to anticipate no interference with the predesigned three-dimensional fold of the *de novo* protein.

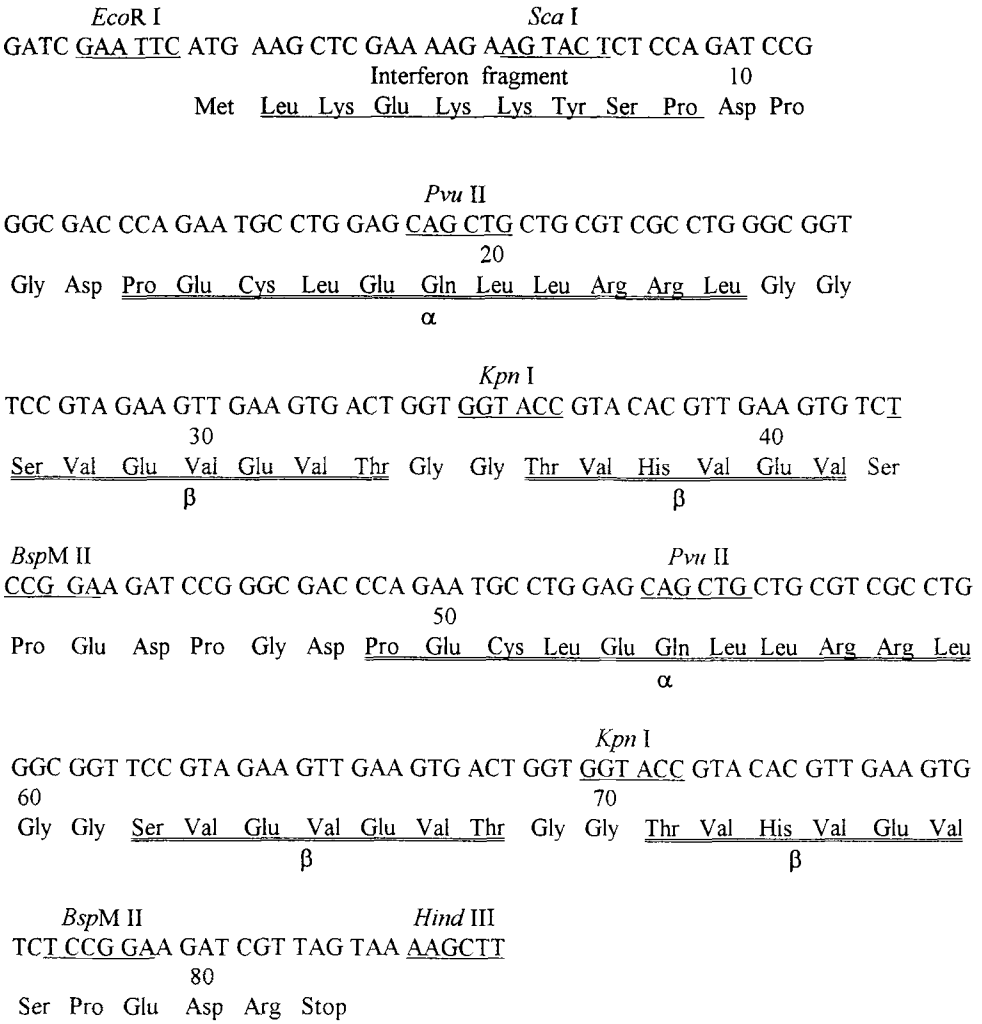

Fig. 3. Amino acid sequence of albebetin incorporating the interferon fragment and nucleotide sequence of the corresponding gene. Amino acids of the inserted interferon fragment, restriction sites, and α-helical and β-structural regions are indicated.

Synthesis of Albebetin with Incorporated Interferon Fragment and its Structural Test

The gene of albebetin incorporating an interferon fragment (Fig. 3) was obtained by introducing the synthesized double-strand oligonucleotide fragment (32 bases) that corresponds to the amino acid sequence LKEKKYSP into the proper restriction sites of the albebetin gene. The interferon fragment was included into albebetin after its first methionine residue. Chimerical albebetin was synthesized in the cell-free expression system as described in Materials and Methods. The polypeptide obtained has a molecular mass of 9 kDa according to the SDS-PAGE, and it is close

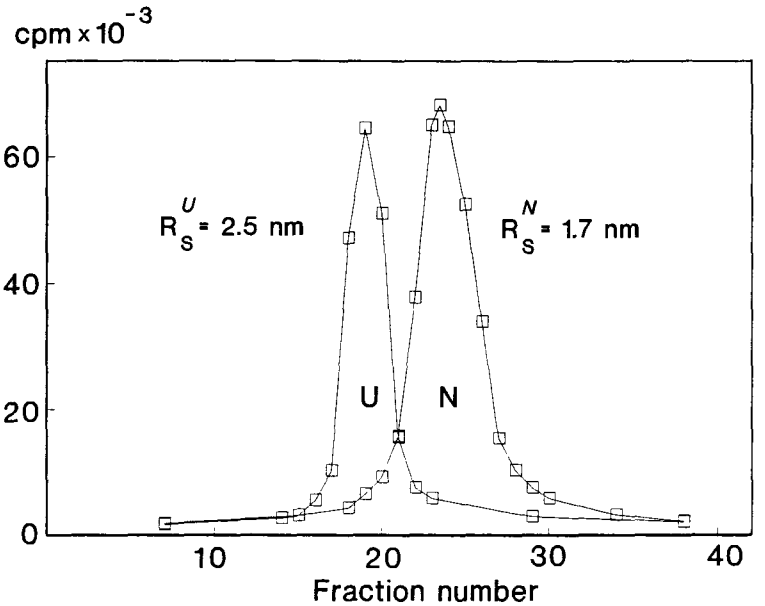

Fig. 4. FPLC chromatography of albebetin with an incorporated interferon fragment in 20 m*M* HEPES, 200 m*M* potassium acetate buffer, pH 7.6 ("native" condition, N), and in the same buffer containing 9*M* urea (U). R_S is the calculated Stocks radii.

to 8.729 kDa calculated from the amino acid content. To check whether the protein is monomeric and compact, we studied it by size-exclusion FPLC under "native" conditions and in the presence of 9*M* urea. Using the calibrated gel-filtration column, the hydrodinamic dimensions (Stokes radius) of the molecule was determined. The data (Fig. 4) show that albebetin incorporating the interferon fragment is a monomeric protein, and its Stokes radius under the "native" conditions is considerably lower than that in 9*M* urea. The comparison of the Stokes radius obtained with those for other globular proteins (*8*) confirms that the chimerical protein is as compact as native globular proteins. It should be noted that its Stokes radius under the "native" conditions does not exceed 1.7 nm for albebetin itself (*11*), i.e., the chimerical protein is at least as compact as albebetin according to this parameter. In 9*M* urea, the Stokes radius of albebetin incorporating the interferon fragment exceeds slightly the value of 2.4 nm reported for albebetin (*11*) in accordance with the molecular mass increase.

To check the stability of the chimerical protein, we studied its unfolding by urea using the FPLC chromatography and urea-gradient electrophoresis. It was shown (*19*) that if the exchange rate between the native and the unfolded states of protein is slow in comparison with the characteristic time of the chromatography, two well-separated elution peaks (corresponding to compact and unfolded protein) are observed at urea concentrations within the transition region. This is the case for urea-

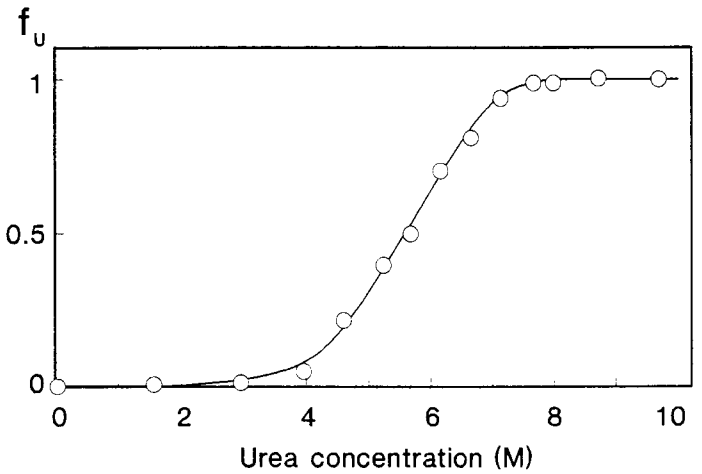

Fig. 5. Urea-induced unfolding of albebetin with the incorporated interferon fragment. F_u is the relative area of the FPLC elution peak that corresponds to the unfolded (non-compact) protein.

induced unfolding of albebetin with the incorporated interferon fragment (data not shown). Figure 5 shows the dependence of the relative area of elution peak, which corresponds to the unfolded molecules on urea concentration. An increase in the relative number of unfolded molecules in the range of urea concentrations 3–7*M* reflects cooperative unfolding of the chimerical protein and confirms that in the absence of urea this protein has a compact structure. The similar result was also obtained by urea-gradient electrophoresis (data not shown). It should be noted that the chimerical protein stability against unfolding by urea is typical of common globular proteins, and the cooperativity of the unfolding transition is also normal for such a molecular mass (*20*). Accordingly, we can assume that the inserted interferon fragment does not significantly disburb the predesigned three-dimensional structure of albebetin.

Biological Activity of Albebetin with Incorporated Interferon Fragment

To testify to the affinity of albebetin with an incorporated interferon fragment to murine thymocyte receptors, we investigated the protein-inhibitory effect on the binding of radiolabeled (^{135}I) octapeptide LKEKKYSP to the receptors. Earlier (*10*) the Scatchard analysis of the binding revealed only one type of receptor for the octapeptide on thymocyte cells (the Scatchard plot represented a straight line). The data shown in Fig. 6 indicate that the chimerical protein competes effectively with the octapeptide for the thymocyte receptor. Using these data, we determined the value of IC_{50} and the inhibition constant (*9*). As follows from Table 1, the chimerical protein has a high affinity to the murine thymocyte receptor and binds it even more effectively than interferon itself. Taking into account a

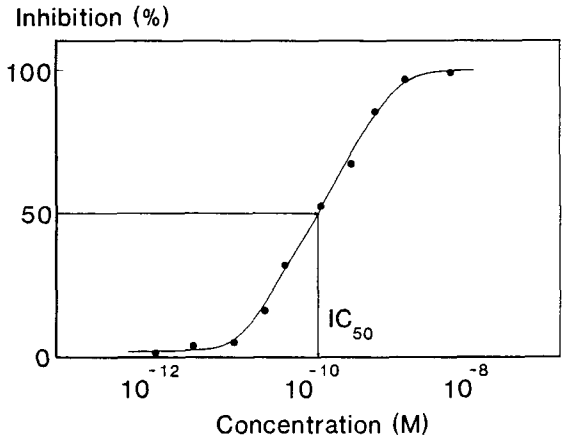

Fig. 6. Inhibitory effect of albebetin with the incorporated interferon fragment on binding of ^{135}I-labeled interferon α_2 octapeptide ($4.4 \times 10^{-11}M$) to the murine thymocyte. The arrow indicates the IC_{50} value corresponding to 50% of inhibition.

Table 1
Parameters of Binding of Albebetin with the Incorporated Interferon Fragment, Human Interferon α_2, and the Octapeptide Interferon Fragment LKEKKYSP to the Thymocyte Receptors

Polypeptide	K_d, M	IC_{50}, M	K_i, M
Albebetin with interferon fragment	Not determined	1.6×10^{-10}	1.4×10^{-11}
Human interferon α_2	2.5×10^{-10}	9.9×10^{-9}	8.6×10^{-10}
Octapeptide LKEKKYSP	4.2×10^{-12}[a]	—	—

[a]Taken from ref. *10*.

very high affinity of the octapeptide *per se* to the receptors (Table 1), it can be assumed that conformational restrictions of the octapeptide in the chimeric protein are lower than those in the interferon molecule.

Tests of the chimerical protein blast-transforming activity confirm its high biological effect on the thymocyte blast-transformation reaction. Figure 7 shows the activation by albebetin with the incorporated interferon fragment in comparison with that by human interferon. One can see that the former induces the thymocyte blast transformation even more efficiently than interferon at concentration as low as $10^{-11}M$.

Further experiments are required for studying structural properties of chimeric protein and its biological activity in more detail. However, the data presented allow us to assume that we did obtain the *de novo* protein with the predesigned structure and biological activity. We think *de novo* proteins may be used as the carriers of various biological activities. This work demonstrated a possibility to transfer a biological function mediated by a short contiguous peptide to the *de novo* protein albebetin. Contiguous

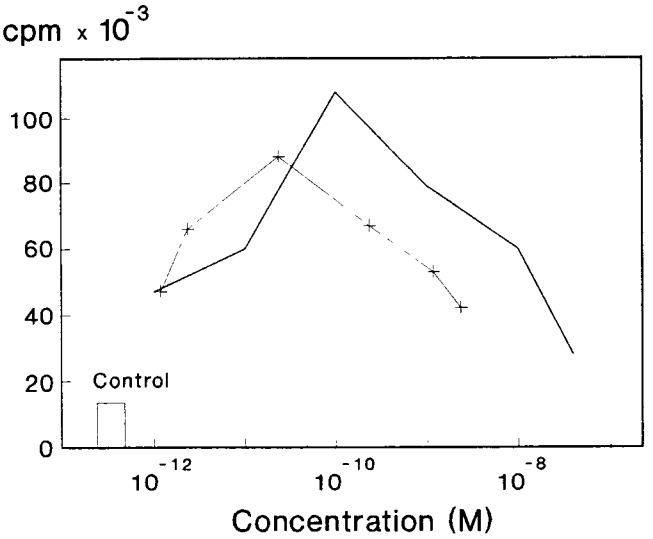

Fig. 7. Activation of the murine thymocyte blast transformation by albebetin with an incorporated interferon fragment (-+-) in comparison with that (—) by human interferon α_2 (*10*). The radioactivity of [^{3}H] thymidine absorbed by the thymocyte cells (counts per minute, cpm) is plotted as the ordinate, and the protein molar concentration is plotted as the abscissa. The control value corresponds to the blast transformation reaction without adding proteins.

amino acid sequences may imitate even such a biological activity that is mediated by the spatially arranged functional sites, and a technique was elaborated capable of searching for such sequences (*21*). It includes incorporation of the random oligopeptides into the surface protein of filamentous phages and the functional sequence selection by their affinity with respect to the specific receptors. This technique expands the sphere of possible biological activities that may be transferred from one protein to another including the *de novo* proteins.

ACKNOWLEDGMENT

A part of this work on the cell-free expression and investigation of radiolabeled proteins was performed by V. V. Chemeris and A. N. Fedorov. G. Nyakatura participated in albebetin expression in *E. coli*, and E. V. Navolotskaya tested the biological activity of the chimeric protein. Expression in *E. coli* and structural testing were performed in the laboratories of H. Bloecker and D. Schomburg. We cordially thank all of them for their valuable contributions. We are grateful to A. V. Finkelstein, O. B. Ptitsyn, and V. P. Zav'yalov for fruitful discussions. This work was made possible in part by grant N1Z000 from the International Science Foundation and a fellowship from the Alexander von Humboldt Foundation for one of us (D. A. D.)

REFERENCES

1. Regan, L. and DeGrado, W. F. (1988), *Science* **241,** 976–978.
2. Sander, C. (1994), *TIBTECH* **12,** 163–167.
3. Fedorov, A. N., Dolgikh, D. A., Chemeris, V. V., Chernov, B. K., Finkelstein, A. V., Schulga, A. A., Alakhov, Yu. B., Kirpichnikov, M. P., and Ptitsyn, O. B. (1992), *J. Mol. Biol.* **225,** 927–931.
4. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982), *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
5. Gurevich, V. V., Pokrovskaya, I. D., Obukhova, T. A., and Zozulya, S. A. (1991), *Anal. Biochem.* **195,** 207–213.
6. Erickson, A. H. and Blobel, G. (1983), *Methods Enzymol.* **96,** 38–49.
7. Ackers, G. L. (1970), *Adv. Protein Chem.* **24,** 343–443.
8. Uversky, V. N. (1993), *Biochemistry* **32,** 13,288–13,298.
9. Cheng, Y. C. and Prusoff, W. H. (1973), *Biochem. Pharmacol.* **22,** 3099–3108.
10. Zav'yalov, V. P., Navolotskaya, E. V., Abramov, V. M., Galaktionov, V. G., Isaev, I. S., Kaurov, O. A., Kozhich, A. T., Mayorov, V. A., Prusakov, A. N., Vasilenko, R. N., and Volodina, E. Yu. (1991), *FEBS Lett.* **278,** 187–189.
11. Chemeris, V. V., Dolgikh, D. A., Fedorov, A. N., Finkelstein, A. V., Kirpichnikov, M. P., Uversky, V. N., and Ptitsyn, O. B. (1994), *Protein Eng.* **7,** 1041–1052.
12. Ptitsyn, O. B. and Finkelstein, A. V. (1980), *Quart. Rev. Biophys.* **13,** 339–386.
13. Ptitsyn, O. B. and Finkelstein, A. V. (1983), *Biopolymers* **22,** 15–25.
14. Finkelstein, A. V. (1983), *Program "ALB" for Protein Polypeptide Secondary Structure Calculation and Prediction*. Deposited at the Brookhaven Protein Data Bank, Upton, NY, at the EMBL Database, Heidelberg, FRG; distributed by request.
15. Finkelstein, A. V., Badretdinov, A. Yu., and Ptitsyn, O. B. (1991), *Proteins: Struct. Funct. Gene.* **10,** 287–299.
16. Provencher, S. W. and Gloeckner, J. (1981), *Biochemistry* **20,** 33–37.
17. Argos, P. (1987), *J. Mol. Biol.* **197,** 331–348.
18. Gabrielian, A. E., Dolgikh, D. A., Ivanova, N. B., and Kirpichnikov, M. P. (1992), in *Proc. 2nd Russian-Israel Symposium on Peptides and Proteins,* Moscow, p. 37.
19. Corbett, R. J. T. and Roche, R. S. (1984), *Biochemistry* **23,** 1888–1894.
20. Uversky, V. N. and Ptitsyn, O. B. (1994), *Biochemistry* **33,** 2782–2791.
21. Wetzel, R. (1991), *Protein Eng.* **4,** 371–374.