Characterization of Globin Domains: Heme Binding to the Central Exon Product

Charles S. Craik; Steven R. Buchman; Sherman Beychok


Stable URL:
http://links.jstor.org/sici?sici=0027-8424%28198003%2977%3A3%3C1384%3ACOGH3E2.0.CO%3B2-E

Proceedings of the National Academy of Sciences of the United States of America is currently published by National Academy of Sciences.

Your use of the JSTOR archive indicates your acceptance of JSTOR’s Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR’s Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/nas.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.
Characterization of globin domains: Heme binding to the central exon product

(Hemoglobin genes/protein folding/evolution)

CHARLES S. CRAIK, STEVEN R. BUCHMAN, AND SHERMAN BEYCHOK

Departments of Chemistry and Biological Sciences, Columbia University, New York, New York 10027

Communicated by Walter Gilbert, December 14, 1979

ABSTRACT We have prepared and isolated the peptide fragments coded for by the three exons of the human β-globin gene, using the arginine-specific protease clostripain (EC 3.4.22.8). The region encoded by the central exon (amino acid residues 31–104) contains an arginine at position 40. This site was less susceptible to cleavage than the two sites that correspond to the exon–intron boundaries, and the isolated central fragment was an approximately equimolar mixture of the entire central fragment, β1–31, 104, and the somewhat smaller fragment contained within it, β1–104. This central fragment mixture bound heme stoichiometrically and tightly at micromolar concentrations, generating a strong Soret absorption band as well as characteristic absorption bands in the visible spectrum. The Soret band occurred at the same wavelength and had the same shape as in hemoglobin, exhibiting an intensity greater than 5% that achieved when native intact β-globin is reconstituted with heme. Nearly the full intensity was regained when an equivalent of heme was added to the unfraktionated digest, suggesting that the noncovalently associated side fragments add precision to the fit of the heme pocket. Three controls were used in establishing the specificity of heme binding to the central fragment mixture. Similar, but preliminary, experiments have also been undertaken with α-globin. A clostripain digest containing α1–31 and α32–141 bound heme, yielding a Soret band identical to that observed in α subunits reconstituted from the native globin chains and heme. Measurements of circular dichroism spectra as indices of secondary structure suggested a role for the side exon products in the acquisition of the native three-dimensional structure of hemoglobin. These experiments confirm a prediction of W. Gilbert that the product of the central exon of the globin gene is a complete functional domain that binds heme tightly and specifically.

Recent investigations of eukaryotic genes have demonstrated that many have mosaic structures, in which expressed sequences of DNA (exons) are separated from others by intervening noncoding sequences (introns) (1–8). These discoveries have raised several important questions: What is the role of the introns? What are the RNA processing mechanisms? What is the relationship, in terms of structure and function, between the protein segment that would be coded for by an exon and the corresponding segment in the intact protein? How do these findings bear on evolutionary processes (9–14)?

Gilbert has formulated a hypothesis to answer some of these questions (9, 10). Among its features are that introns serve multiple purposes, including that of providing DNA space between exons, thereby increasing recombination frequency between exons in different transcription units. If each exon codes for a protein of particular function that has been selected for, then joining of the protein leads to a new protein of improved or multiple function. The hypothesis has a number of consequences that have been discussed by Gilbert (10). For the present report, we focus on one, namely, that the protein segments coded for by exons are functional domains.

During the past 2 years, it has been discovered that distinct exons of the heavy and light chain immunoglobulin genes code for each functional element and structural domain within the molecule (5, 6, 15, 16). This precise exon–function correlation has provided a stimulus for research to determine whether the correlation is general in eukaryotic proteins. Because the genes coding for several hemoglobins have been extensively investigated and the protein structures are well known, these proteins are ideal candidates for studies along the same lines. Globin genes so far cloned and sequenced consist of five segments, three exons separated by two introns (2–4, 17). In terms of the β-globin sequence, the three exons code for residues 1–30, 31–104, and 105–146, respectively. The corresponding positions in α-globin are 1–31, 32–99, and 100–141.

In contrast to the immunoglobulins, however, the junctions occur within long α-helical segments, and there are no clear structural features that would suggest demarcation into functional domains. Indeed, conformational studies on heme-free chains and on various peptide fragments derived from myoglobin (18) and hemoglobin (19, 20) prior to the elucidation of gene structure have tended to emphasize loss of structure after removal of heme and chemical or enzymatic cleavage. However, Gilbert (9), Blake (14), and Argos and Rossman (21) have noted that the product of the central exon contains the proximal heme-binding histidine as well as most of the residues that make contact with heme (22). The Gilbert hypothesis predicts that the central exon product should bind heme specifically and tightly.

As an extension of our work on structure/function determinants in α- and β-globin chains (23–26), we have performed tests of heme binding to the central exon product. Our results confirm the prediction that the central fragment of the globin chains is a heme-binding domain. They also shed light on other aspects of the relationship between the exon-coded fragments and the structure of the globin chains.

MATERIALS AND METHODS

All chemicals used were highest purity grade. Hemoglobin, separated subunits, heme-free globin chains, and heme dicynide solutions were prepared as described (25). Absorption and circular dichroism spectroscopy of heme-free solutions and reconstituted mixtures were performed at 4 °C (25).

Protolyis. Clostripain (27, 28) (clostridiopeptidase B, EC 3.4.22.8; Worthington) was activated at a concentration of 1 mg/ml in 2 mM dithiothreitol/2 mM CaCl2. Digestions were performed at room temperature in distilled deionized water containing 2 mM CaCl2 and 2 mM 2-mercaptoethanol. Protein concentrations ranged from 15 to 30 mg/ml and the substrate-to-enzyme (wt/wt) ratio was 150:1. The time of hy-

Abbreviations: α As a superscript to an α or β indicates that the chain has no heme attached; δ indicates that the chain has heme bound; dansyl, dimethylaminophthalene-5-sulfonyl.
drolysis ranged from 10 to 60 minutes. Digestions were stopped by making the EDTA concentration of the reaction cocktail 10 mM, lowering the pH to 2 with formic acid, or both. Peptic digestion was accomplished at 37°C in 4.5% (wt/wt) formic acid, pH 2.0. Pepsin (Sigma) was incubated with globin at a substrate-to-enzyme (wt/wt) ratio of 100:1 for 60 min. Immediate lyophilization of the digestion mixture was employed to remove formic acid and to stop the reaction.

Amino-Terminal Analysis. Dimethylaminonaphthalene-5-sulfonyl (dansyl) chloride, micropolyamide layer plates, and dansylated standards were purchased from Pierce. Aminoterminal residues of peptides were identified as the dansyl derivatives, according to procedures of Gray (29) and of Woods and Wang (30).

RESULTS
Preparation and Characterization of the Fragments. The arginine distribution in the α and β subunits is illustrated in Fig. 1. Although the heme-containing subunits can be cleaved by clostripain, they are poor substrates (reaction times of 8–48 hrs) in comparison to the heme-free globin chains. A 10-min clostripain digest of β was analyzed on standard NaDodSO4/polyacrylamide gels consists of one diffuse band (apparent molecular weight of 12,000) and a high molecular weight aggregate that does not enter the gel. These associated fragments were resolved by using NaDodSO4/urea/polyacrylamide gels. A time course of the clostripain digestion of β reveals that the arginyl cleavage of β is nearly complete after 60 min. To determine the amount of undigested globin chain remaining, a 60-min digest was fractionated on a Sephadex G-75 column in 8 M urea/1 M propionic acid. Less than 8% of intact globin (molecular weight 16,000) remained.

The first evidence that a heme-binding core exists in a clostripain digestion mixture was found when one equivalent of hemin dicyanide was added to a β digest and the mixture was fractionated on a NaDodSO4/urea/polyacrylamide gel and stained with a heme-specific reagent. Heme staining was evident in bands corresponding to aggregates of undigested globin, as well as bands corresponding to molecular weights of 16,000 and 9000.

The various arginyl peptide fragments of β were separated by gel filtration on a Sephadex G-50 column as described in Fig. 2. The distribution of tryptophan and tyrosine residues in β globin is such that relative concentrations of the fragments cannot be judged by A280 values. Fractions were therefore initially pooled not only with respect to A280 but also according to their molecular weights (see Fig. 2). After lyophilization, these fractions were analyzed on NaDodSO4/urea/polyacrylamide gels calibrated with the same molecular weight markers [1000–20,000 (31)] that were used to calibrate the Sephadex G-50 column. The results are shown in Fig. 3 and tabulated in Table 1. Included in Table 1 are the results of the dansyl chloride end-group analysis of these fractions. Amino acid composition data and tyrosine-to-tryptophan ratios determined by spectrophotometry (32) provided additional evidence for identification of the peptides.

Cyanoheme Addition Studies. The β31–104, 41–104 mixture

---

* 3,3′-Dimethoxybenzidine/barium peroxide (1:4, wt/wt) in 50% acetic acid freshly prepared. Stain for 10 minutes, then rinse gel with distilled H2O.

---

Fig. 1. Clostripain cleavage sites of α and β globin. Solid arrows indicate labile arginine sites. Dashed arrows denote arginine sites less prone to cleavage. The numbers below the line refer to the sequence number of the arginine residues. For β globin, the solid arrows demarcate the exon-coded peptides.

Fig. 2. Gel filtration of clostripain digestion products of β globin. A 10-min clostripain digestion of 80 mg of β globin was filtered through a column (15.5 × 120 cm) of Sephadex G-50 (fine) in 9% (wt/wt) formic acid at 6 ml/hr, monitored at 280 nm, and collected in 1.7-ml fractions. Fractions A–G were pooled separately as indicated, lyophilized, and analyzed. See Table 1 and Fig. 3 for peptide characterization. V0, exclusion volume.

Fig. 3. NaDodSO4/urea/polyacrylamide gel of α′ clostripain digest, β clostripain digest, and gel filtration chromatography products of β digest (see Fig. 2). The fragments of the globin sequence are shown on the sides; BPB, bromphenol blue. The gel was prepared as described (31) with modifications. The separation gel was 12.5% polyacrylamide with a bisacrylamide-to-acrylamide wt ratio of 1:10. The gel buffer was 0.2 M Tris-H3PO4, pH 6.8. A 10% stacking gel with a bisacrylamide-to-acrylamide ratio of 1:15 was employed, and its buffer was the same as the reservoir buffer, 40 mM Tris-H3PO4, pH 6.3. Lane 1, fraction G; lane 2, fraction F; lane 3, fraction E; lane 4, fraction D; lane 5, fraction C; lane 6, fraction B; lane 7, fraction A minus 2-mercaptoethanol (note loss of bands); lane 8, fraction A; lane 9, 10-min β clostripain digestion; lane 10, same as lane 9 but three times the concentration (note β31–40 band); lane 11, same as lane 9; lane 12, 15-min α′ clostripain digestion; lane 13, α′. Material with molecular weight greater than 16,000 (β1–140) probably represents aggregates arising from oxidation of the unprotected sulfhydryls. All lanes were heavily overloaded (40 μg per lane) to reveal the presence of trace peptides. Identification of β31–140, β31–146, β1–104, α′29–141, and α′3–31 was based solely on estimated molecular weights derived from known standards.
was found to contain 51% $\beta^a_{31-104}$ and 49% $\beta^b_{31-104}$ from the $A_{280}$ value of a known dry weight of the mixture. Molar extinction coefficients of the separated fragments were similarly calculated according to their amino acid composition and dry weight.

Table 2 lists extinction coefficients and wavelength maxima of Soret bands when hemin dicyanide was added in equimolar amounts to the various chains, digestion mixtures, and fragments listed. Included in Table 2, also, are results using two controls, a 1-hr peptic digest of $\beta^b$ globin, and bovine serum albumin, which forms a strong complex with heme (33).

The Soret and visible region spectra of the untreated $\beta^b$ globin itself, of the central fragment mixture, and of the controls after addition of one equivalent of hemin dicyanide are shown in Fig. 4. In Fig. 4A, the spectrum of reconstituted $\beta^b$ globin is compared to that of hemin dicyanide. With the exception of the absolute value of the extinction coefficient at 420 nm (see Table 2, and below), the central fragment mixture, $\beta^b_{31-104}$, 41-104, shown in Fig. 4B, and the intact native $\beta^b$ globin exhibit virtually identical spectral characteristics upon combination with one equivalent of hemin dicyanide. The Soret maxima of both protein spectra coincide at 420 nm; the widths at half-height of the Soret bands are the same, and the characteristic visible band at 540 nm of the cyanomet complex is also evident in Fig. 4B.

The controls are compared to hemin dicyanide in Fig. 4C, in which it is seen that the peptic digest of $\beta$ globin shows a spectrum that is very similar to that of free hemin dicyanide. The cyanometalbumin spectrum departs substantially from free hemin dicyanide, especially in the red shift of the Soret peak. However, the maximum is at 413 nm, rather than 420 nm.

### Table 1. Characterization of globin chains and fragments

<table>
<thead>
<tr>
<th>Protein or peptide</th>
<th>$M_r$, amino acid composition</th>
<th>Approx. $M_r$</th>
<th>NH2-terminal amino acid</th>
<th>Dansyl end</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta^b$</td>
<td>15,950 Val</td>
<td>16,000</td>
<td>Val</td>
<td>16,000</td>
</tr>
<tr>
<td>$\beta^b_{31-104}$</td>
<td>12,110 Leu, Val, Phe</td>
<td>13,200</td>
<td>Leu</td>
<td>2300</td>
</tr>
<tr>
<td>$\beta^b_{31-104}$</td>
<td>7,510 Leu, Phe</td>
<td>8,200</td>
<td>Leu</td>
<td>1900</td>
</tr>
<tr>
<td>$\beta^b_{1-104}$</td>
<td>4,500 Leu, Val</td>
<td>10,600</td>
<td>Leu</td>
<td>1900</td>
</tr>
<tr>
<td>$\alpha^a$</td>
<td>15,130 Val</td>
<td>16,000</td>
<td>Val</td>
<td>16,000</td>
</tr>
<tr>
<td>$\alpha^a_{32-141}$</td>
<td>12,940 Val</td>
<td>13,500</td>
<td>Val</td>
<td>13,500</td>
</tr>
<tr>
<td>$\alpha^a_{31-31}$</td>
<td>3,190 Met</td>
<td>4,000</td>
<td>Met</td>
<td>4,000</td>
</tr>
</tbody>
</table>

* From gel chromatography and NaDodSO4/urea/polyacrylamide gel electrophoresis.
† Average molecular weight.

### Table 2. Soret band absorption parameters

<table>
<thead>
<tr>
<th>Protein or peptide</th>
<th>$\varepsilon_{420}$, m$^{-1}$ cm$^{-1}$</th>
<th>$\lambda_{max}$, nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta^b$</td>
<td>105.0</td>
<td>420</td>
</tr>
<tr>
<td>$\beta^b + h^*$</td>
<td>86.0</td>
<td>420</td>
</tr>
<tr>
<td>$\beta^b$ digest + h</td>
<td>79.0</td>
<td>420</td>
</tr>
<tr>
<td>$\beta^b$ digest + $\alpha^b + h^*$</td>
<td>78.0</td>
<td>420</td>
</tr>
<tr>
<td>$\beta^b_{31-104}$, 41-104 + h</td>
<td>62.0</td>
<td>420</td>
</tr>
<tr>
<td>Bovine serum albumin + h</td>
<td>44.0</td>
<td>413</td>
</tr>
<tr>
<td>$\beta^b$ peptic digest + h</td>
<td>26.0</td>
<td>395 (broad)</td>
</tr>
<tr>
<td>$\beta^b_{1-30}$ + h</td>
<td>19.0</td>
<td>395</td>
</tr>
<tr>
<td>$\beta^b_{31-40}$ + h</td>
<td>15.9</td>
<td>396</td>
</tr>
<tr>
<td>h</td>
<td>18.6</td>
<td>395</td>
</tr>
<tr>
<td>$\alpha^b$</td>
<td>118.0</td>
<td>418</td>
</tr>
<tr>
<td>$\alpha^a + h^*$</td>
<td>84.0</td>
<td>418</td>
</tr>
<tr>
<td>$\alpha^a$ digest + h</td>
<td>83.0</td>
<td>418</td>
</tr>
</tbody>
</table>

All samples contained 20 mM potassium phosphate buffer/1 mM EDTA (pH 5.7) and were at 4°C. h refers to hemin dicyanide prepared as described. Globin digest contains less than 8% undigested globin (see text).
* Reconstitution of $\alpha$ and $\beta$ globins with hemin dicyanide to give the same Soret extinction coefficients as those of the untreated subunits, $\alpha^a$ and $\beta^b$, has been achieved many times in our laboratory. However, to do this requires conditions of pH and ionic strength that were not routinely employed in this study as well as purification of the reconstituted subunits by chromatography, dialysis, and other procedures, as reported elsewhere (25).
† The extinction coefficient reported is only due to the contribution of the $\beta^b$ digest. The extinction coefficient of the equimolar mixture of $\alpha^b$, $\beta^b$ digest, and h is 93 m$^{-1}$ cm$^{-1}$ and should be compared to the value of 120 m$^{-1}$ cm$^{-1}$ for hemoglobin (see footnote immediately above).
‡ The spectrum of hemin dicyanide in 20 mM potassium phosphate/1 mM EDTA (pH 5.7) also has a $\lambda_{max}$ at 365 nm. However, this peak changes primarily in intensity, not position, in the presence of heme-binding protein.
DISCUSSION

An important feature of the hypothesis that exon translation products represent functional elements within a final protein is that such functional elements in different combinations with other exon products, and after additional mutation, may serve in a variety of roles while preserving an underlying functional similarity. Therefore, in searching for correspondence between the attributes of an isolated fragment presumed to be a domain and the exon-coded segment of the intact larger protein, it may be critical to strike a balance in expectation between the general function of a domain believed to be common to many proteins and the specific one found in one or a group of closely related proteins.

We have selected the criterion of heme binding as a general function of the central exon product, with the specification that the binding should be tight and stoichiometric. As a more specific criterion appropriate for the hemoglobins, we have selected the intensity, position, and shape of the Soret band of the metmyoglobin derivatives. Spectral characteristics in the near ultraviolet and visible regions have long been used to characterize cytochromes, for example, and to distinguish native hemoglobins from their denatured forms (35).

The central fragment mixture, upon addition of one equivalent of hemin dicyanide, gives a Soret band at the same wavelength maximum as hemoglobin itself and with somewhat greater than 70% the extinction coefficient observed when heme is added to undigested ϒ globin. With the digest itself, the extinction coefficient is within 10% of that of ϒ globin. It should be noted, also, that the central fragment mixture gives an absorption band at 540 nm, which is characteristic of the metmyoglobin derivative of all hemoglobins. Moreover, although the slopes of the plots in Fig. 5, after equivalence, are steeper than expected, probably due to some weak heme binding of contaminating fragments, the absence of curvature as equivalence is approached in this concentration range indicates that the binding constants are greater than $10^6$ M$^{-1}$.

The results leave no doubt that both fragments $e^{31-104}$ and $e^{31-104}$ bind heme tightly and with specificity, comparable to hemoglobin chains, in terms of spectral features. Thus, the principal prediction of the Gilbert hypothesis (9, 10) for this case is confirmed. In view of the increased extinction coefficient of the digest as compared to the central fragment, it appears likely that the structural features defining the heme pocket are induced by heme binding to the central fragment, but may be made more precise when the side fragments are present, even if not covalently joined. This last conclusion, however, cannot be drawn with certainty until all the constituent fragments are purified and reconstituted.

Although this work demonstrates tight and stoichiometric binding of heme to the central exon product, several experiments must be done to give further definition to the specificity of the binding and to establish the relationship between the gene element represented by the exon and the expressed protein segment. In addition to purifying all fragments and reconstituting the isolated fragments, it is necessary to understand heme binding to fragments that contain some of the heme contact residues and the proximal histidine but that are not the central exon products exactly.

Finally, experiments should be conducted to see whether the central heme-binding fragment binds oxygen reversibly. Furthermore, the distinctive features of oxygenation in hemoglobin are closely coupled with heme–heme interaction (or cooperativity), which requires subunit interfaces and many detailed intersubunit interactions. Eaton has pointed out that the central exon in the ϒ globin gene also codes for the residues involved in the $\alpha_1$–$\beta_2$ interface, and that the major interactions...
of the $\alpha_1$-$\beta_1$ interface of hemoglobin are in the side exon fragments (36). This suggests that studies are necessary of the interactions of each of the exon-coded fragments in each chain with the other chain and its domains.

We gratefully acknowledge the patience and skill of Dr. Mollie Pflumm, who provided invaluable support from the work's conception. We also thank Dr. Jeremy Luchins, Ms. Yvonne Swiderska, and Ms. Jyoti Misra for helpful discussions, and Ms. Rhea McDonald for her preparation of the manuscript and assistance with the figures. S.B. is indebted to Dr. J. E. Darnell, Jr., for an invaluable discussion of several of the concepts underlying this work. This investigation was supported in part by Grant PCM 77-08537 awarded by the National Science Foundation and Grant CA 13014 awarded by the National Cancer Institute, DHEW. C.S.C. is supported in part by Public Health Service Training Grant GM 07216.