Characterization and Crystallization of Core Streptavidin*

(Received for publication, April 20, 1987)

Arno Pähler‡, Wayne A. Hendrickson‡, Mary Ann Gawinowicz Kolks§, Carlos E. Argaránall, and Charles R. Cantor‡

From the ‡Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics, and the Departments of §Medicine and ¶Genetics and Development, College of Physicians and Surgeons, Columbia University, New York, New York 10032

We have characterized a streptavidin product that had been reduced to a minimal size that still retained full biotin-binding activity. This core streptavidin is proteolyzed at both ends at points that correspond closely with the termini of hen egg white avidin. Core streptavidin is more soluble than is the parent molecule. We have grown three different types of crystals of core streptavidin. The symmetry properties of these crystals prove that the molecule is a tetramer organized in tetrahedral (D2) point symmetry. The crystallographic response to the interaction of biotin with core streptavidin indicates that some conformational change accompanies ligand binding. We are attempting to determine the three-dimensional structure of streptavidin and its complex with selenobiotin from these crystals of core streptavidin.

Streptavidin is an isolate from Streptomyces avidinii that takes its name from this bacterial source and from its similarity to hen egg white avidin in binding affinity for biotin (1). Despite very different origins, the properties shared by avidin and streptavidin point to evolutionary relatedness. The most conspicuous feature shared by these proteins is their remarkable affinity for binding biotin, affinity that is unprecedented for noncovalent interactions (2). Similarities in physical properties are also striking. Both molecules are approximately the same size and both are tetramers comprising identical subunits (1, 2). Most significantly, the sequence of streptavidin as recently determined from the cloned DNA (3) is clearly homologous with the amino acid sequence found for avidin (4). However, some notable differences do distinguish the two proteins. Avidin is a glycoprotein with a single carbohydrate moiety per subunit and it possesses a disulfide bridge as well as 2 methionines. Streptavidin, on the other hand, has no carbohydrate and is completely devoid of sulfur-containing residues.

Much interest attends the extreme affinity of these proteins for biotin. The dissociation constants for both proteins are on the order of 10−15 M and the complex with one biotin per subunit is extremely stable over a wide range of temperature and pH. Although the biotin-binding activities of avidin and streptavidin have been known for more than 40 and 20 years, respectively (1, 5), a conclusive biological function has not been assigned to this remarkable interaction nor has an explanation for the high affinity been established. Binding studies with biotin and more than two dozen analogs have been informative but do not reveal the binding mechanism in detail. The imidazolidone portion of the biotin molecule must be intact for strong binding, but all parts contribute to the binding. On the protein side, 3 or 4 tryptophan residues appear to interact directly with bound biotin and a lysine residue is implicated as being nearby (2). Biotin binding to avidin is noncooperative, but the binding to isolated subunits is less strong than it is to whole molecules (6).

Streptavidin and avidin are more than exotic digressions of nature. In recent years, their unique properties have been exploited to devise powerful and widely applicable tools for microbiology, biochemistry, and medical science. Applications include immunoassays, hybridization assays, lymphocyte activation, immunotherapy, localization of antigen, affinity chromatography, and others that have recently been reviewed (7–9). In some applications, nonspecific interactions of the carbohydrate moiety of avidin interfere with the desired biotin-specific binding; thus, streptavidin, which has no carbohydrate, offers some distinct advantages for biotechnology. We have undertaken a crystallographic analysis of the streptavidin structure in order to investigate the nature of its tight complex with biotin. Since the cloned DNA is available, the mechanism for binding can be examined by site-directed mutagenesis once the binding site is characterized. We also expect that the structure might suggest designs for other useful intentional mutants or fusion products. There have been three reports of avidin crystallizations (10–12), but problems in reproducibility have until now thwarted structure determination. Difficulties in avidin crystallization might possibly arise from heterogeneity in the carbohydrate portion (12, 13). Streptavidin is free of this complication.

RESULTS

Protein Characterization

The streptavidin that we have crystallized was supplied as a gift from Apcel Ltd. (Berkshire, United Kingdom). In contrast to earlier reports of low solubility (1) and our own experience with material obtained from another commercial source, this preparation of streptavidin is highly soluble in water, streptavidin concentrations of at least 10% (w/v) are obtained without any detectable insoluble material. Assays of the material were provided by the manufacturer. SDS1-poly-

1 The abbreviations used are: SDS, sodium dodecyl sulfate; MPD, 2-methyl-2,4-pentanediol; PEG, polyethylene glycol; AS, ammonium sulfate.
acrylamide gel electrophoresis at an acrylamide concentration of 15% (w/v) showed a single, sharp band. Biotin binding, assayed as for avidin (14), ranged from 13.8 to 17.2 μg of biotin bound per mg of dry sample weight in five different analyses. This corresponds to from 0.75 to 0.93 biotin molecules per subunit assuming that the sample consists entirely of streptavidin and that its molecular weight is as determined here (13,200/chain). We conclude that these preparations are essentially pure and nearly free of biotin.

Streptavidin is known to be susceptible to proteolysis near the N terminus (3) and the Apel product has been processed by an undisclosed protocol to a minimal size that still retains full biotin-binding activity. We have characterized this material and find that oligopeptides have been removed from both ends of the initial gene product. Accordingly, we use the name "core streptavidin" to describe the Apel-processed protein that we have crystallized.

Core streptavidin was subjected to 15 cycles of Edman degradation and automated gas-phase sequence analysis. Yield from the first cycle indicates that essentially all of the sample was sequenced and the resulting analysis revealed two N-terminal sequences, approximately 62% of the material corresponds exactly to the sequence of holo streptavidin starting from Ala-13 and the remainder has the streptavidin sequence from Glu-14 onward. The amino acid composition of core streptavidin was also determined. Analyses were made after 24 and 48 h of acid hydrolysis with two samples taken at each time point. Quantities were normalized to the average of His and Phe of 2.0 residues. Thr and Ser values were adjusted according to ribonuclease A standards run under identical conditions. The Gly value was adjusted for a high glycine concentration (0.9 mol/mol of protein) in the dissolved lyophilized sample. Averages from the four samples as thus adjusted give a composition of 4.2 Lys, 2.0 His, 3.8 Arg, 11.6 Phe, 0.2 Cys, 6.8 Val, 0.0 Met, 2.9 Ile, 7.1 Leu, 5.8 Tyr, and 2.0 Phe. Trp was not determined. It is clear from these results that truncation has occurred at the C terminus as well as at Asx, 19.5 Thr, 11.1 Ser, 8.3 Glx, 2.3 Pro, 16.3 Gly, 19.3 Ala, 0.4 Ser at 5 min to 0.41 Ala and 0.12 Ser at 30 min. Carboxypeptidase Y immediately released these same amino acids (0.45 Ala, and 0.57 Ser at 2 min) and then others began to appear at 5 min. The composition of released products at 30 min (1.0 Ala, 1.02 Ser, 0.21 Lys, 0.20 Thr, 0.12 Val, 0.12 Pro, 0.10 Asp, 0.10 Phe, and 0.07 Leu) continued this pattern. Alanine and serine release was continuing to rise at 30 min with both enzymes.

These carboxypeptidase data are consistent with a mixture having termini at both Ser-139 and Ala-138. Assuming that digestion at 30 min is limited by availability of unfolded substrate and by enzyme specificity (very slow or negligible attack of carboxypeptidase A on the Pro-Ser bond (15) and slow carboxypeptidase Y cleavage of the Lys-Pro bond (16)), the proportions of the two components can be estimated from the alanine to serine ratios. By this analysis, the carboxyl termination of core streptavidin is 70% at Ser-139 (59% from the A data and 82% from the Y data) and 30% at Ala-138. Alternatives, such as an Ala-137 terminus, cannot be rigorously excluded; but this would conflict with the protein composition and magnitude of alanine release. The relative compositions of lysine, threonine, valine, proline, phenylalanine, and aspartic acid released at 30 min by carboxypeptidase Y are as expected given the poor reactivity of the Gly-His bond (16) that precedes the C-terminal sequence of Asp-Thr-Phe-Thr-Lys-Val-Lys-Pro-Ser-Ala-Ala-Ser.

Our characterization of core streptavidin is summarized in Fig. 1.

Crystallization

We carried out crystallization experiments using the hanging drop variant of the vapor diffusion technique. In a typical experiment, a 5- or 10-μl droplet of protein solution is placed on a siliconized microscope coverslip and mixed with an equal volume from a reservoir of 0.5 or 1 ml of precipitant solution in one of 24 wells of a Linbro tissue culture tray. The coverslip is then inverted over this reservoir well until an air tight grease seal. Our experiments were carried out in an incubator held at 20 °C.

Lyophilized core streptavidin was dissolved in distilled water to concentrations of 10-40 mg/ml. The precipitating agents used in our studies included ethanol, 2-propanol, 2-methyl-2,4-pentanediol (MPD), polyethylene glycols 3350, 4000, and 6000 (PEG), ammonium sulfate (AS), ammonium nitrate, sodium phosphate, sodium sulfate, and sodium chloride. These chemicals were used as purchased from various commercial suppliers without further purification. Crystals were obtained with all of these precipitants except for ethanol and 2-propanol. However, in many cases only microcrystalline precipitates or very small isolated crystals were obtained.

---

* B. Ellis and G. Humphreys, personal communication.
Crystalline Streptavidin

Crystals suitable for x-ray studies were obtained from AS and MPD. We have characterized three basically different crystalline modifications of core streptavidin.

Tetragonal Type A Crystals—These crystals, the first to be obtained, grew as square plates from AS, PEG, and ammonium nitrate at slightly acidic pH. The crystals from PEG were too small for diffraction studies but had the same morphology as larger ones grown from the ammonium salts. Good crystals were obtained from a protein solution at a starting concentration of 10–20 mg/ml in 50 mM potassium acetate and 200 mM sodium chloride, adjusted to pH 4.5, after vapor~equilibration against a reservoir of 30% saturated AS in the same buffer. Crystals usually formed after 12–18 h, but they sometimes appeared as early as 2 h after setting up the experiment. Growth terminates after about 2 days, and the largest crystals measure 0.5 × 0.3 × 0.1-mm. Conditions for reproducibly growing large tetragonal crystals have not been established.

Precession photographs taken with CuKα radiation from a rotating anode source were used to establish, from the Laue symmetry and systematic absences, that these crystals belong to the body-centered tetragonal space group 14/22. The diffraction pattern extend to 2.5-Å spacings on still photographs from rather small crystals. Precise unit cell parameters (Table I) were determined by refinement against the angular positions of a set of reflections measured on a Rigaku AFC-5 four-circle diffractometer. The crystal habit of type A crystals is determined by plate faces of the form [001]. These are bounded by [110] edge faces to form squares or additionally by a [210] face to produce a right triangular plate.

Monoclinic Type B Crystals—Another crystalline modification was obtained from MPD and also from PEG or AS. MPD produces exclusively these type B crystals. These crystals can be grown reproducibly by dissolving core streptavidin in 40% MPD in water (v/v) to obtain a protein concentration of 30–40 mg/ml and then equilibrating a hanging drop of this solution against 50% MPD. Final crystal size depends critically on the reservoir MPD concentration. In order to control the delivery of the viscous MPD liquid, these solutions were mixed on a weight rather than volume basis. Crystals usually appear after 1–2 days and they reach their full size after several more days. The largest type B crystals that we have obtained measured 15 × 0.6 × 0.2-mm. An often encountered problem is the intergrowth of otherwise single crystals. These can usually be separated, but single crystals can also be grown directly with the addition of slight amounts of β-ocetylglucoside. However, we did not find the effect to be as quantifiable or predictable as described elsewhere (17).

Diffraction patterns of type B crystals recorded by precession photography have the symmetry and systematic absences of space group P212121. Lattice parameters determined from diffractometer measurements from a particular crystal are reported in Table I. These typical lattice constants also pertain to crystals grown in the presence of β-ocetylglucoside. However, occasional polymorphs with distinctly different unit cells have also been observed. One crystal of type B2 (a = 58.5A, b = 79.1A, c = 46.8A, β = 98.0°) and another of type B3 (a = 58.5A, b = 76.3A, c = 44.6A, β = 98.7°) have been characterized. Variables controlling this polymorphism are not understood since variants are sometimes found together with standard B1 crystals in the same crystallization drop. The morphology of type B crystals is somewhat variable. Initial crystals from AS were featherlike objects elongated along c. Well-formed crystals are rectangular rods bounded by [100] faces and slightly wider [010] faces along the direction of elongation with blunt truncation by a (001) face at one end and, often, by a more rounded or multifaced termination at the other end.

Measureable diffracted intensities were observed to 2.0 Å spacings in zonal data collected from a type B1 crystal on the AFC5 diffractometer. A complete data set was measured to 3.0 Å spacings from the same crystal. Radiation damage was moderate. The self rotation function was computed from type B diffraction data and the diad section is displayed in Fig. 2a.

We have attempted to soak monoclinic crystals in d-biotin and to crystallize the biotinylated protein. Soaking of biotin into crystals causes cracks along the long axis and co-crystallization of core streptavidin and biotin gives amorphous precipitates under conditions that otherwise produce type B crystals.

Ortborhombic Type C Crystals—These crystals grow from ammonium sulfate solutions at lower pH than is optimal for tetragonal crystals. A typical experiment that yields crystals suitable for x-ray studies consists of vapor diffusion from a reservoir of 30–32% saturated AS in 0.1 M potassium acetate at pH 4.0 into a hanging drop made up of 5 μl from the reservoir and 5 μl of core streptavidin dissolved in water to a concentration of 30 mg/ml. They appear as diamond-shaped plates within 48 h and reach full size, as large as 1.0 × 0.6 × 0.2-mm, within a few days. Crystals can be obtained at any pH from 2.5 to 5.0. At the high end of the pH range it appears, based on morphology, that orthorhombic, tetragonal, and monoclinic crystals can coexist in the same droplet. Crystalization attempts with biotin present do not initially yield crystals under the same conditions that readily crystallize core streptavidin alone; however, small but chunky crystals (about 0.1 mm) were found after several months in one experiment. Soaking with d-biotin or d-selenobiotin produces superficial cracks along the long axis of the diamond plates and also renders the crystals more fragile.

The space group of type C crystals was determined from diffractometer measurements of 5-Å data in the appropriate primitive, orthorhombic unit cell. These measurements and checks of potential symmetry equivalents showed the body centered systematic absences and mirror symmetries characteristic of space groups I222 and I212121, among the allowed.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Complex</th>
<th>Space group</th>
<th>a (Å)</th>
<th>b (Å)</th>
<th>c (Å)</th>
<th>β (°)</th>
<th>V (Å³)</th>
<th>Zₙ</th>
<th>Solvent fraction %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type A</td>
<td>Native</td>
<td>14,22</td>
<td>58.11</td>
<td>86.18</td>
<td>58.49</td>
<td>(2)</td>
<td>46.80</td>
<td>(1)</td>
<td>98.85 * (2)</td>
</tr>
<tr>
<td>Type B1</td>
<td>Native</td>
<td>22,22</td>
<td>96.07</td>
<td>105.24</td>
<td>98.74</td>
<td>(2)</td>
<td>47.16</td>
<td>(2)</td>
<td>98.74 * (2)</td>
</tr>
<tr>
<td>Type B2</td>
<td>Native</td>
<td>22,22</td>
<td>96.07</td>
<td>105.24</td>
<td>98.74</td>
<td>(2)</td>
<td>47.16</td>
<td>(2)</td>
<td>98.74 * (2)</td>
</tr>
<tr>
<td>Type B3</td>
<td>Native</td>
<td>22,22</td>
<td>96.07</td>
<td>105.24</td>
<td>98.74</td>
<td>(2)</td>
<td>47.16</td>
<td>(2)</td>
<td>98.74 * (2)</td>
</tr>
</tbody>
</table>

Table 1

Crystalllographic data on core streptavidin polymorphs

Standard deviations in unit cell parameters are given in parentheses and refer to the last cited digit. Zₙ designates the number of protein subunits per crystallographic asymmetric unit. Solvent fraction numbers assume a molecular mass of 13,200 daltons/subunit and the computed partial specific volume of 0.711 ml/g.
Fig. 2. Self-rotation functions of diad symmetry in core streptavidin crystals. Rotation functions were calculated with the fast rotation function (24) using data in the range from 10.0 to 5.0-Å spacings and a Patterson cutoff radius of 30 Å. Data are presented as zenithal equiarea projections ($\psi$ is displayed with radial position proportional to $\sin \psi/2$ and $\psi$ advances counterclockwise from zero at the c axis) for the $\chi = 180^\circ$ sections. Circles are drawn corresponding to 90°, 60°, and 30° and lines of constant $\psi$ are drawn at 45° intervals. The maps are contoured at intervals of 0.5 $u$ starting at 1.5 $u$ where $u$ is the root mean square value of the entire rotation function. Density features that specify the location of molecular 2-fold symmetry axes are indicated by A, B, and C. a, monoclinic type B1 crystal. b, orthorhombic type C1 crystal.

possibilities for handed molecules. Lattice parameters for native type C1 crystals and for type C2 crystals soaked with biotin were also determined from diffractometer measurements, and these are reported in Table I. Type C crystals soaked in selenobiotin also show shortening of the c axis as for biotin binding. The degree of shortening is somewhat variable for both biotin and selenobiotin. Reflections from type C crystals are observable to beyond 2-Å spacings in still photographs. The correspondence between diffraction and morphology establishes that the crystal habit is determined by plates of the form [100] and edges of the form [011].

Diffraction data to 3.0-Å spacings have been measured at four wavelengths near the absorption edge of the selenium in a type C2 crystal. Resistance to radiation damage is good since this crystal showed diffraction to beyond 2.4-Å spacings after these synchrotron measurements and collection of another 3-Å data set at home on the diffractometer. The multiwavelength anomalous diffraction data from these crystals have been analyzed to give the selenium positions. This analysis breaks the space group ambiguity and indicates that type C crystals belong to I222 rather than I212121. The diad symmetry section of a self rotation function for type C crystal is shown in Fig. 2b.

**DISCUSSION**

The relationship of core streptavidin to the initial gene product and to hen egg white avidin is shown in Fig. 1. It is striking that at the sites of cleavage at both ends of streptavidin align closely with the avidin termini, and that these sites include the same sequence of Ser-Ala-Ala. Both cuts could conceivably be made by the same proteolytic enzyme. The heterogeneity observed in these cleavages is similar to that seen when other proteases digest proteins with adjacent cleavage sites, i.e. endopeptidase activity may be so low that second, trimming cuts are not made (18). Accessibility of these cleavage sites to proteolysis suggest that the N- and C-terminal portions of native streptavidin are flexible. On the other hand, the inaccessibility of the C terminus in native core streptavidin to carboxypeptidase suggests that this structure is compactly folded. The amino acid composition of core streptavidin is remarkably similar to that reported by Green (19) for streptavidin and is distinct from that found by Argaraiia et al. (3). As noted before (3), truncated streptavidin might unwittingly have been studied in earlier work. It is also evident that the termini of core streptavidin are very similar to those of avidin. With allowance for eight short insertions and deletions, 33% of the residues of core streptavidin are identical to those of avidin.

The heterogeneity that we observe in core streptavidin most probably derives from uncorrelated double cuts at both ends giving rise to four chains: 14-138, 13-138, 14-139, and 13-139. Thus, core streptavidin is reduced from the initial gene product of 159 residues, which has a molecular mass of 16,450 daltons, to a mixture of 125, 126, and 127 residue chains ranging from 13,096 to 13,254 daltons. This hypothesis completely explains our amino- and carboxy-terminal sequence data and the single peak found in SDS-polyacrylamide gel electrophoresis. There is no evidence for internal proteolytic nicks that might have survived the Apcel purification. We have not attempted to determine the heterogeneity of core streptavidin from dissolved crystals. However, it does not seem likely that the crystals have fractionated the mixture. All three types of crystals can grow rapidly and prolifically from the same batch of protein. Moreover, there is ample...
precedent for the incorporation of sequence heterogeneity or terminal cleavage heterogeneity into crystal lattices as is seen in the structures of crambin (20) and ubiquitin (21) refined at high resolution.

The molecular size and crystallographic data permit us to deduce the unit cell contents and the molecular symmetry. Core streptavidin chains have a weighted average molecular mass of 13,200 daltons and a calculated partial specific volume of 0.71 ml/g. From these data and the unit cell volume for the tetragonal crystal, it is clear that the asymmetric unit of type A crystals must contain a single subunit. If there were two, the solvent content would be unacceptably low at 15%. Since streptavidin is known to be tetrameric (1,3), the only possible molecular location in these crystals is at a special position in 1422 with 222-point symmetry. This proves that core streptavidin has Da molecular symmetry. Since there are no diad axes in space group P21, these crystals must have a multiple of whole molecules in the asymmetric unit and volume considerations dictate that only one can be present. The rotation function results identify the locations of the molecular diads in type B crystals and these are mutually orthogonal within 0.2° as expected from Da symmetry. In the case of type C crystals the rotation function results show that one diad axis of the molecule coincides with the crystallographic diad along a and that the other diads are noncrystallographic. Thus, in this case, there are two subunits in each crystallographic asymmetric unit.

The binding of biotin imparts a marked change in the crystallization behavior of streptavidin and the soaking of biotin into crystals causes cracking and cell constant changes. Similar effects are seen with avidin. For example, the P2221 crystal form (11) undergoes a change from a = 71.6 to a = 74.7 Å upon soaking in biotin (note that Ref. 11 incorrectly attributed the published cell constants to native avidin instead of the biotin complex). Apparently, some conformational change accompanies ligand binding to avidin and streptavidin. These affect intermolecular contacts but they are sufficiently mild that, in favorable cases, they can be accommodated in the lattice even if non-isomorphously. We are not aware of previous studies on the binding of selenobiotin to avidin or streptavidin, but the similarity that we observe between biotin and selenobiotin complexes is consistent with the isomorphism of biotin and selenobiotin crystal structures (22).

We are actively pursuing the structure analysis of core streptavidin in the type C and type B crystals. We have had no success in binding heavy atoms to these crystals, although trials have been limited. However, analysis of the selenobiotin complex in type C2 crystals by use of multiwavelength anomalous diffraction is proceeding well. We anticipate that this structure will reveal the nature of biotin binding and define the subunit interfaces. This will provide a basis for planning intentional mutants of the cloned DNA (3) to probe these interactions. We anticipate that streptavidin will be of the "all β" class of protein structures. Raman spectroscopy of avidin indicates a secondary structure with 55% of residues in β strands and 10% in α-helices. The similarity of these values to the secondary structure found in the x-ray analysis of prealbumin (45% β and 9% α) has been noted previously (20). Prealbumin (now called transthyretin after its function) is also a tetramer composed of small (128 residues) subunits, but no substantial homology exists between it and either avidin or streptavidin. An early suggestion of similarities between avidin and lysozyme (19) is also unsupported by sequence comparisons.

Acknowledgments—We thank Apel Ltd., United Kingdom, for supplying us with the streptavidin preparation used in these experiments, Dr. A. Marquet (Université Pierre et Marie Curie, Paris) for a supply of selenobiotin, M. Maslowska and M. Schneider (Free University, Berlin) for their help in the first successful crystallization experiments, Drs. S. Birken and K. Drickamer for discussions about the protein characterization, and Drs. G. Humphreys, R. M. Buckingham, and B. Ellis (Apel) for discussions about core streptavidin.

Note Added in Proof—The structure of selenobiotinyl streptavidin (crystal Type C2) has now been solved. After stereochromically restrained refinement with bond ideality of 0.027 Å, the agreement index is R = 0.29 at 3.1 Å resolution.

REFERENCES

A. Pähler, unpublished results.