Proc. Natl. Acad. Sci. USA
Vol. 77, No. 6, pp. 3393–3397, June 1980
Biophysics

Model for haptoglobin heavy chain based upon structural homology

(three-dimensional structure/serine proteases/protein–protein interaction)

JONATHAN GREER

Department of Biological Sciences, Columbia University, New York, New York 10027

Communicated by Cyrus Levinthal, January 25, 1980

ABSTRACT **A model has been constructed for haptoglobin heavy chain by using the known sequence homology to the mammalian serine proteases. The three-dimensional structures for three serine proteases, chymotrypsin, trypsin, and elastase, were compared and the structural features that are conserved in all three were extracted. The haptoglobin heavy chain sequence was aligned to the sequences of the three serine proteases by maximizing sequence homology in the regions of conserved structure. The resulting alignment shows that haptoglobin heavy chain must be very closely homologous to these proteases in structure as well as in sequence. Coordinates were derived for the heavy chain by using the homologous structures. The problems associated with these coordinates are outlined and methods for solving them are indicated. The features of the haptoglobin heavy chain structure are described. Implications of the structure for the very strong interaction between this subunit and hemoglobin are discussed.**

Haptoglobin (Hp) is a serum glycoprotein that is present in many mammalian species (1). Its function is to form a strong and stable complex with hemoglobin that has been released from erythrocytes and foster the recycling of heme iron. Hp is a tetramer composed of two light and two heavy chains. A variety of experiments indicate that it is the Hp heavy chains (HpH) that bind a hemoglobin $\alpha\beta$ dimer during complex formation (2, 3). Kurosky *et al.* (4–6) have reported that the sequence of HpH is clearly homologous to the mammalian serine protease family. The possibility exists, therefore, of determining the tertiary structure of HpH by fitting the sequence into the known structure for the serine proteases (7,8).

Browne *et al.* (9) first used comparative model building from the known structure of lysozyme to predict the structure of the homologous α-lactalbumin. McLachlan and Shotton (10) applied this technique with mixed success (11) to construct α-lytic protease from the structure of elastase (12). Their model suffered from the fact that the sequence homology between the bacterial and mammalian enzymes is extremely weak.

We report here the application of comparative model building to HpH. The strategy adopted was, first, to analyze the structural features common to the known serine proteases. The clear sequence homology between HpH and the other serine proteases was used to align the HpH sequence in such a way as to maximize its agreement with the structural characteristics of the serine proteases. Atomic coordinates were then constructed based upon this alignment.

METHOD OF BUILDING COORDINATES

Mammalian serine proteases

Atomic resolution crystal structures are available for three mammalian serine proteases: chymotrypsin (13, 14), trypsin (15, 16), and elastase (12). Atomic coordinates for chymotrypsin (13), trypsin (15), and elastase (12) were obtained from the *Atlas of Macromolecular Structure* (17).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

The coordinates for trypsin and elastase were rotated and translated into the same coordinate frame as chymotrypsin, and the three homologous structures were then compared (Fig. 1). Large parts of the three structures were found to be closely equivalent; that is, the respective α carbons lie within 1.0 Å of each other. In other regions of the molecule, the three structures differ quite dramatically. Upon examination, the closely equivalent regions correspond to the two β barrels that form the structural core of these serine proteases, as well as the large α helix in the molecule. The regions that deviate from each other correspond to the loops and turns and lie on the surface of the molecule (see Fig. 1). It is entirely reasonable that structural deviation should occur on the surface of the molecule whereas the core structure remains strongly conserved.

By use of this structural comparison (Fig. 1), a residue-by-residue correspondence can be established between the three structures (Fig. 2). The regions that are structurally closely equivalent in all three proteins are enclosed in boxes in this figure; these structurally equivalent regions also show strong sequence homology. Equivalent structural sites would be expected to have identical or closely homologous side chains. The structural homology breaks down in the external loops, where the three structures deviate (outside of boxes in Fig. 2). Sequence homology is almost completely absent from these loops. In addition, all the additions and deletions between the three protein sequences lie outside the boxes in the external loops.

HpH sequence

The HpH sequence was kindly provided to us by A. Kurosky prior to publication (18). If the HpH structure is consistent with the structural motifs of the serine proteases, good sequence homology should be observed in the closely equivalent structural regions (boxes of Fig. 2). Thus, the criterion for aligning the HpH sequence was to search for stretches of residues that were either identical to or structurally similar to the sequences in the boxed regions (see legend to Fig. 2). Minimum base change was not used as the criterion for alignment because it was not the purpose of this work to determine how the HpH sequence arose in evolutionary history (19–21). Rather, the purpose was to construct an atomic model and, therefore, structural equivalence with the consequent conservation of sequence were of prime importance.

Fig. 2 shows the derived homology for the HpH sequence. Sequences of chymotrypsin, trypsin, and elastase that are homologous to those in HpH are shown in capital letters. A good stretch of homologous sequence can be found for every boxed region of the molecule. Thus, the alignment in these areas is unambiguous. However, in the looped regions, sequence homology is absent and the alignment is more arbitrary. Nevertheless, it is true that all additions and deletions in the HpH sequence relative to the other three homologues occur in the external loops, usually where there were already deviations in chain length among the three homologues.

Thus, it appears that the clear sequence homology of HpH,

Abbreviations: Hp, haptoglobin; HpH, haptoglobin heavy chains.

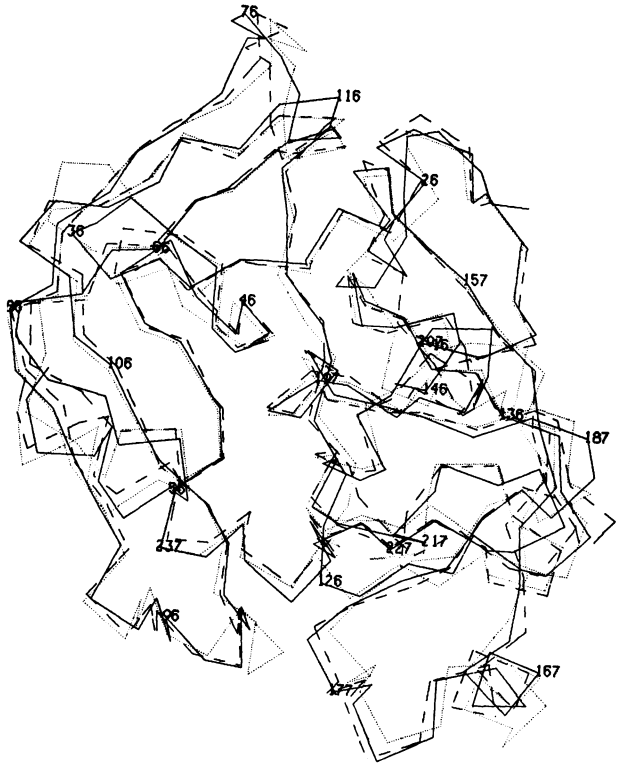

FIG. 1. α-Carbon plots of chymotrypsin (—), trypsin (- - -), and elastase (·····). Residue numbers are those of chymotrypsin. All three molecules were placed in the chymotrypsin coordinate frame. The transformation for trypsin to chymotrypsin was obtained by aligning the molecules initially on a graphics cathode ray tube terminal in order to assign the residue correspondences between the two structures. A least-squares fit was performed on the α carbons. All α carbons that deviated by more than 1.5 Å were deleted and the least-squares fit was performed again. The final root mean square deviation for 147 α carbons was 0.632 Å. For elastase, the transformation to the chymotrypsin coordinate system was calculated by a least-squares fit with the same 147 α carbons, giving a root mean square deviation of 0.854 Å. The coordinate transformations for the above two proteins are:

Trypsin to chymotrypsin

$$\begin{pmatrix} 0.582 & 0.257 & 0.771 \\ -0.664 & 0.697 & 0.270 \\ -0.469 & -0.669 & 0.577 \end{pmatrix} \begin{pmatrix} -24.461 \\ -45.336 \\ 51.068 \end{pmatrix}$$

Elastase to chymotrypsin

$$\begin{pmatrix} -0.821 & -0.268 & 0.504 \\ 0.368 & -0.923 & 0.107 \\ 0.436 & 0.274 & 0.857 \end{pmatrix} \begin{pmatrix} -4.274 \\ 21.807 \\ -28.818 \end{pmatrix}$$

which was reported by Kurosky *et al.* (4–6), is reflected in a strong structural homology. The HpH sequence fits the structural motif of the serine protease family remarkably well, much more so than do the microbial serine proteases (10, 11).

Construction of coordinates for HpH

The rules for constructing the HpH sequence into the structure of a homologue were as follows. When a matching residue existed in HpH and the homologue, the main chain coordinates of the homologue were used directly. If the side chains were identical, then the side chain coordinates were used directly. If the side chains differed, coordinates for a standard side chain were placed onto the main chain by using the observed χ_1 (if available) of the corresponding residue in the homologue. When a deletion occurred in HpH, the coordinates for the residue in the homologue were skipped. For additional residues in HpH, dummy coordinates were introduced at this stage.

By use of these rules, three versions of the HpH structure were constructed, one from each of the homologues. These can be compared as summarized in Table 1. The structures agree closely in the boxed regions whereas the loops differ considerably. Where all three homologues predict closely equivalent coordinates in main chain and in side chain, then the derived HpH coordinates are likely to be correct. Where the homologues predict differing coordinates for the main chain or for the side chain, it is clear that this region requires careful model building.

Detailed examination of the three homologous sequences and that of HpH suggests that in various parts of the structure, different homologues are more useful models for the HpH structure. For example, the HpH loop at residues 182–186 (C185–C187)* has the same number of residues as that in trypsin. On the other hand, the local main chain at position 110 (C129) in chymotrypsin would accommodate a Tyr residue, whereas the main chain of the half-cystine of trypsin would not. In regions of differing sequence length, wherever possible, the appropriate homologue would be chosen so that the HpH sequence has the smallest relative deletion and not an addition; e.g., the loop at residues 21–22 (C36–38) from chymotrypsin.

From a variety of such considerations a hybrid set of coordinates was compiled from the three homologous HpH structures. The source homologue for each residue is indicated by italicizing the respective residue name in Fig. 2. The transition point between two homologous structures [e.g., 21–22 (C36–C38) or 68–69 (C81–C82)] creates a discontinuity in the chain. In order to simplify model building, they were positioned at an already existing discontinuity (e.g., 21–22) or in a region of close equivalence of the main chain between the homologues (e.g., 68–69) where the discontinuity would be very small. Because of the choice of different homologues for the loop regions, only two regions, 47–50 and 204–205, required building main chain and side chain *ab initio*.

Problems in the constructed coordinates

The coordinates were refined by using a model building and energy minimization program, PAKGGRAF, written by Levinthal and coworkers (22). A detailed description of the coordinate refinement will appear elsewhere. The types of problems encountered will be discussed here.

Overlapping Residues. The method of constructing the coordinates directly from the main chain of the homologues virtually assures that no overlap of main chain will occur in the HpH coordinates. However, in several places, the substitution of side chains results in overlap. In most of the occurrences, the close contacts could be relieved by rotation of one or both of the side chains about the Cα-Cβ bond. In the remaining cases, the regions involved were external loops where modification of the main chain was required in any case (see below).

Additions and Deletions in Loop Regions. The major problem in the derivation of atomic coordinates by homology is construction of the external loops that contain the added and deleted residues (Fig. 2). In most cases, by proper choice of homologous structure the HpH sequence can be treated as having a relative deletion. In two cases, at residues 47–50 and 204–205, additional residues occur in HpH. Starting coordinates for these were generated by model building, taking care that no overlap of atoms resulted. The main chain ϕ and ψ and side chain χ angles of the residues in each loop were set as variables and the loop was closed with a spring closing potential (23). In many of the loops, closing the discontinuity was straightforward

* All residue numbers in this paper have the following convention. Plain residue numbers refer to the HpH sequence. They correspond to the line labeled "HNO" in Fig. 2. Residue numbers preceded by a "C" which appear on the line labeled "CNO" in Fig. 2, refer to the standard chymotrypsinogen residue numbering. Fig. 2 can be used to convert from one nomenclature to the other.

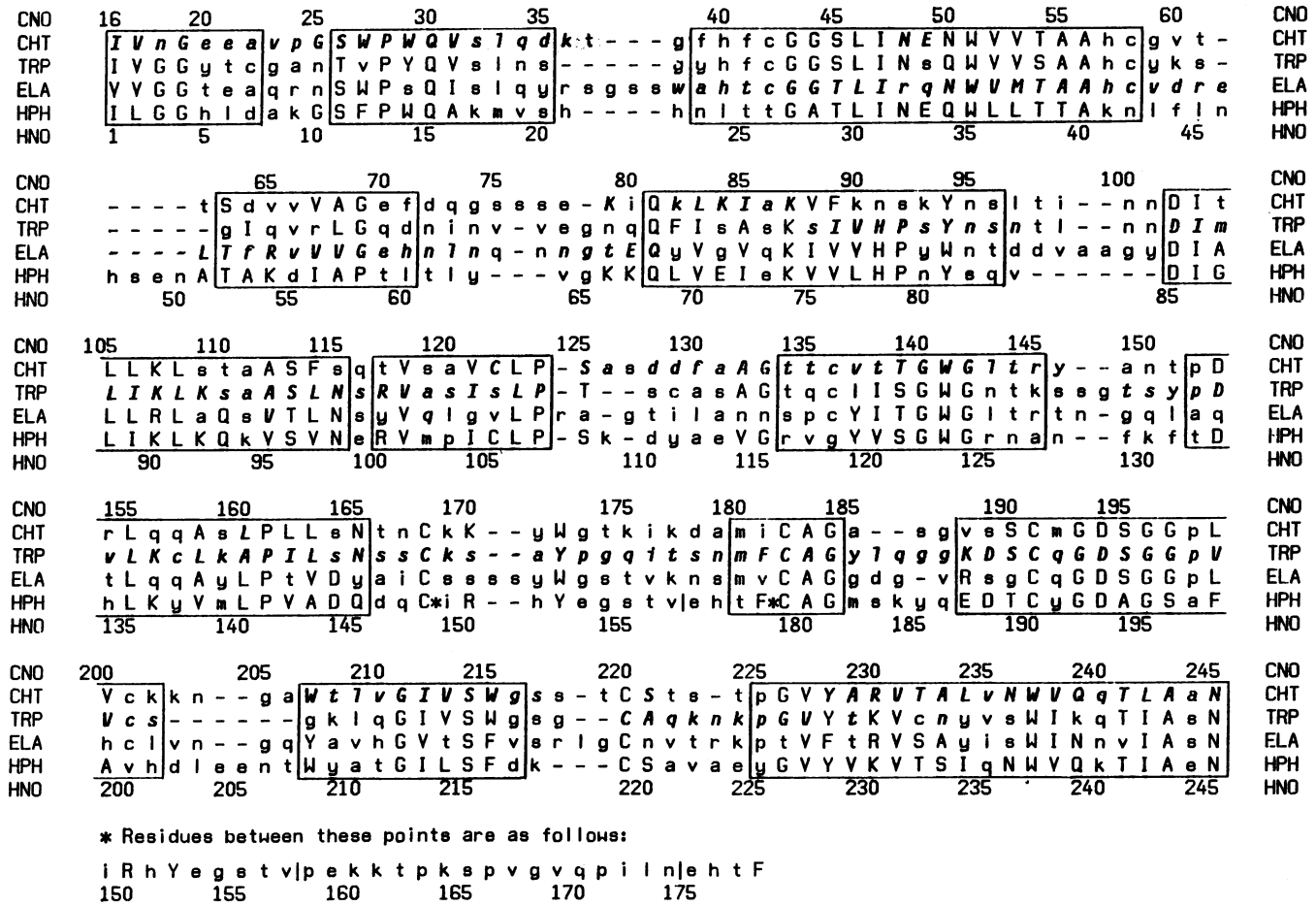

FIG. 2. Sequences of chymotrypsin (CHT), trypsin (TRP), and elastase (ELA) are presented here. They have been aligned by comparison of the three-dimensional structures (see text). The residue numbers on the top row, labeled CNO, are the standard chymotrypsinogen numbering and are preceded by a "C" when used in the text. The boxes indicate residues that are closely equivalent in their structural positions between these three structures (see Fig. 1). Residues outside the boxes deviate significantly in at least one of the three structures. The one-letter code for amino acid residues is: a, A = Ala; c, C = Cys; d, D = Asp; e, E = Glu; f, F = Phe; g, G = Gly; h, H = His; i, I = Ile; k, K = Lys; l, L = Leu; m, M = Met; n, N = Asn; p, P = Pro; q, Q = Gln; r, R = Arg; s, S = Ser; t, T = Thr; v, V = Val; w, W = Trp; y, Y = Tyr. The bottom sequence is the HpH sequence of Kurosky and coworkers (18); the residue numbers for this sequence appear on the bottom line labeled HNO. The sequence was aligned to the others by maximizing sequence homology in the boxed regions of the structure (see text). Those residues in the three serine proteases that are homologous to the HpH sequence are shown in capital letters. Sequence homology is defined as residues that are either identical or structurally similar. Hence, a Ser and Thr are taken to be homologous as is a Val with an Ala or a Leu or an Ile. A Glu or Asp can be homologous to a Lys or an Arg if the residue is pointing out of the molecule into the solvent. On the other hand, a Phe is not normally considered homologous to an Ala or even a Val because it is structurally much larger. In general, more than one residue in a row must be homologous for the homology to be accepted unless such a residue lies adjacent to a homologous stretch of residues. Because of the relative rarity of Trp residues and their unusually large size which gives them a somewhat special role as a spacer in protein structure, the coincidence of Trp residues in the alignment is taken as significant even if it is isolated from other homologous residues. A hybrid molecule was constructed from parts of the HpH structures that were built from each of these three homologues (see text). The homologue that was used as the model to build a particular HpH residue is represented by having its residue name in italics. Stretches of italicized residues mean that main chain and side chain were taken from this homologue. Single italicized residue names mean that just the side chain χ_1 value was taken from this homologue.

because the loop conformation is a turn that can easily accommodate one or two residues more or less (e.g., C36–C38). In several loops, the topology of the molecule is more complex and extensive energy minimization will be necessary to determine the correct conformation (e.g., C72–C80).

One further deviation in the HpH structure is in the methionine loop, which is 13–15 residues long in the homologues (C169–C181) but is 30 residues long in HpH (148–177) (18). Fitting these residues is a major problem. In the present structure, HpH residues were assigned to each of the 13 residues in the trypsin loop so that the additional 17 residues would lie between C177 and C178. The present assignment is quite arbitrary because no good sequence homology was evident in this loop. A further discussion of this loop appears later.

Buried Charged Residues. In order to test for buried charged groups that are energetically very unfavorable, a surface accessibilty calculation (24, 25) was performed by using a program written by Bruce Bush. Six charged residues appear to be buried. Glu-113 is inaccessible to solvent but lies close to the surface and forms a salt bridge with Arg-116. Two residues, Asp-7 and Asp-217, are close to the surface and require only minor changes in the side chain and main chain conformation to become accessible. Asp-85 (C102) and Asp-193 (C194), which are buried in the serine proteases, also appear to be buried in this structure. The last residue is Asp-55, which is pointing deeply into the hydrophobic core of the first β barrel and has no countercharge. Energetically, this is not reasonable. Therefore, this residue is really an Asn or this region of the molecule deviates considerably from the homologous serine protease structures or the sequence homology in this region must be modified. Because residue 55 lies in the middle of a closely equivalent region that shows good sequence homology [52–60 (C63–C71), see Fig. 2] and is surrounded by two such regions [11–20 (C26–C33) and 68–70 (C81–C83)], a large

Table 1. Root mean square (RMS) values for the differences between the three HpH coordinate sets

	Chymotrypsin-HpH vs. trypsin-HpH		Chymotrypsin-HpH vs. elastase-HpH		Trypsin-HpH vs. elastase-HpH	
	No.	RMS, Å	No.	RMS, Å	No.	RMS, Å
All atoms						
Boxed	1254	1.72	1254	1.61	1254	1.77
Loops	404	4.00	447	3.73	412	3.66
Both	1658	2.48	1701	2.36	1666	2.38
Main chain						
Boxed	644	0.85	644	0.93	644	0.83
Loops	200	2.38	224	2.62	208	2.33
Both	844	1.38	868	1.55	852	1.35
Side chain						
Boxed	610	2.31	610	2.09	610	2.39
Loops	204	5.12	223	4.58	204	4.64
Both	814	3.25	833	2.97	814	3.11

structural deviation would be surprising. The buried charge can be removed by a local realignment of the sequence homology so that residues 53–61, rather than 52–60, correspond to C63–C71 of Fig. 2. This homology is poorer than that shown in Fig. 2; however, if residue 55 is truly an Asp, it would be structurally much more reasonable.

PROPERTIES OF THE MODEL HPH STRUCTURE

Structural homology of HpH

The pattern of close sequence homology of HpH to the mammalian serine proteases demonstrates quite definitively the very close structural homology of HpH to the serine proteases (Fig. 2). The structural homology permits the confirmation of a variety of features of the HpH structure first pointed out by Kurosky *et al.* (4) based upon sequence homology. The two disulfide bridges in HpH, 148–179 (C168–C182) and 190–219 (C191–C220), correspond to disulfide bridges that are found in all three homologues (Fig. 3). In addition, Cys-105 (C122), which joins HpH to the Hp light chain (26), has its counterpart in chymotrypsin, where Cys C122 forms a disulfide bridge with the A peptide.

Kurosky *et al.* (4–6) have suggested that the active-site residues of the serine proteases, Ser C195, His C57, and Asp C102, are Ala, Lys, and Asp in HpH, respectively. In addition, the Asp that precedes the reactive Ser is conserved. This residue forms an internal salt bridge with the α-amino group of Ile C16 in chymotrypsin, trypsin, and elastase. The close homology of the NH_2 terminus (Fig. 2) and the presence of this Asp suggest that this buried salt bridge may be formed in HpH (4) (Fig. 3).

Substrate specificity in the serine proteases is bestowed by the residue at position C189, a Ser in chymotrypsin and elastase and an Asp in trypsin. This residue is an Asp in HpH, as pointed out by Kurosky *et al.* (6), suggesting trypsin-like specificity. The restriction of the elastase substrate specificity to small side chains is believed to be due to blockage of the hydrophobic pocket that binds the side chain of the substrate by Val C216 and Thr C226, which replace glycines found in chymotrypsin and trypsin (12). In HpH, these residues are Asp and Gly, respectively (Fig. 3). Therefore, it is not clear to what extent the hydrophobic pocket is blocked. It is true that the Asp at position C189 must be accessible to solvent; hence the pocket must at least be accessible to water. Detailed surface accessibility calculations (25, 27) will be needed to determine the extent of the hydrophobic pocket in HpH.

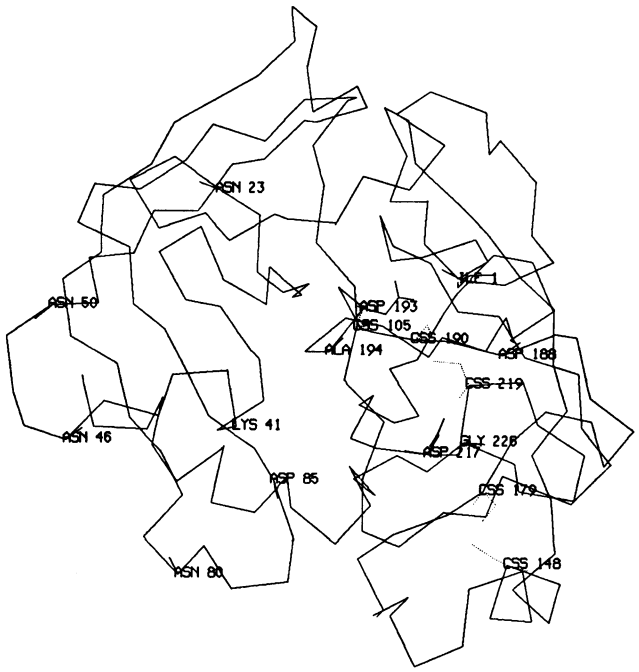

FIG. 3. α-Carbon plot of the HpH structure. The residues that are labeled include the active-site residues, Lys-41 (C57), Asp-85 (C102), and Ala-194 (C195); the specificity residue, Asp-188 (C189); the disulfide bridges at 148–179 (C168–C182) and 190–219 (C191–C220); and Cys-105 (C122), which attaches to the Hp light chain. The salt bridge between Asp-193 (C194) and Ile-1 (C16) is shown. The four Asn residues that bind carbohydrate are labeled 23, 46, 50, and 80. The residues that line the specificity side chain binding pocket, Asp-217 (C216) and Gly-226 (C226), are also labeled.

HpH has approximately 20% carbohydrate bound at four points: Asn-23, 46, 50, and 80 (18). The HpH structure (Fig. 3) shows that all four Asn residues are exposed to solvent and occur in external loops where carbohydrate can easily bind.

Structural features of the HpH model

The sequence comparison and the overlap free construction of HpH indicate the close structural homology of the core parts of this protein. The equivalence is so good that most of the structure does not require further discussion at this time. Even most of the external loops, where additions and deletions of residues occur, fit the pattern of structural features of the serine proteases. However, certain loops appear to differ dramatically from their homologous counterparts. One such example is the loop at residues 61–67 (C72–C80). Close examination of this region in the three homologous structures (Fig. 1) shows that their conformations differ considerably even though they all have the same number of residues in the loop. The HpH sequence in this loop is shorter than those of the other homologues.

The most challenging part of building the model of the HpH structure is the enlarged methionine loop. As already noted, the HpH sequence has 30 residues in this loop rather than the usual 13–15. It is interesting to compare this part of the structure with that of the microbial serine proteases. Delbaere *et al.* (11, 28) have shown that in *Streptomyces griseus* protease A, the Asp C102 loop is diminished by five residues. Instead, the methionine loop is rearranged so that a 13-residue section lies across the Asp loop (Fig. 4). In HpH, the Asp C102 loop is shortened by four residues. Thus, the microbial enzyme may be a good model for the structure of this loop in HpH. In addition, the different conformation of the Met loop in *S. griseus* protease A might suggest an approximate path for the additional 17

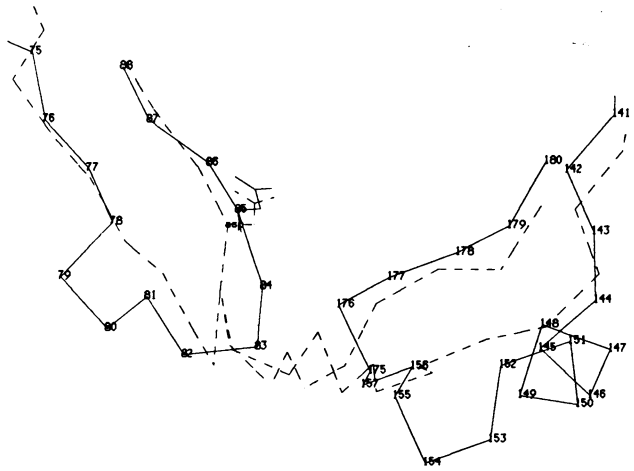

FIG. 4. Comparison of the HpH structure (—) with that of the microbial serine protease *S. griseus* protease A (- - -) in the region of the Asp C102 loop (residues C88–C106) and the methionine loop (C163–C182). The coordinates for *S. griseus* protease A were obtained from the Protein Data Bank at Brookhaven (29). See text for discussion of these structures.

residues of HpH. The main chain of the loop in HpH would follow that of the mammalian serine proteases, as shown in Fig. 4, until the break around residues C177–C178 (157 and 175), at which point the remaining 17 residues would follow the *S. griseus* protease A structure in an antiparallel β sheet heading toward the Asp C102 loop. For the present, the 17 additional residues have been omitted from the structure.

The function of Hp is to bind very strongly to hemoglobin. This binding is via the HpH subunit alone (2). Much evidence suggests that strong protein–protein contacts involve the interaction of large hydrophobic complementary surfaces on the protein with the consequent exclusion of solvent molecules (30). Subunit interactions involving salt bridges have been observed in one of the subunit contacts in tobacco mosaic virus protein disks (31). It is useful, therefore, to examine the surface properties of the constructed HpH coordinates.

If one looks at the charge distribution on the surface of the subunit, only one region of the surface appears to be devoid of charged groups. This region lies immediately adjacent to Cys-105 (C122), which forms the disulfide bridge to the light chain and probably corresponds to the Hp light chain binding site. In all other parts of the subunit surface, there appear to be charged groups distributed quite evenly. The $\alpha_1\beta_2$ surface of hemoglobin is a strong candidate for the interaction site on hemoglobin (32, 33). This surface is also quite polar and contains several charged groups which are involved in the functional properties of hemoglobin (34). It seems likely, therefore, that the HpH–hemoglobin contact involves salt-bridge interactions together with buried hydrophobic surfaces.

The availability of HpH coordinates together with the known crystallographic coordinates for hemoglobin dimers (35) makes it possible to try modeling the interaction between these two molecules. Methods are being developed by using a surface representation of the molecules (27) to search for the correct contact interaction. It is hoped that these studies will lead to the detailed identification of the interaction sites on HpH and on hemoglobin that are responsible for the formation of this exceedingly strong protein complex.

I am indebted to Dr. Alexander Kurosky for making his sequence information on HpH available to me prior to publication. I thank Dr. Barry Honig for helpful discussions. I also thank Dr. Bruce Bush, Noel Kropf, Christos Tountas, Boris Klebansky, and David Yarmush of the Columbia Biology Department Computer Graphics Facility for programs used in the determination and the display of these results. This research was supported by National Institutes of Health Grant HL-16601 and Facility Grant RR-00442 and by the Columbia University Computer Center.

1. Putnam, F. W. (1975) in *The Plasma Proteins*, ed. Putnam, F. W. (Academic, New York), 2nd Ed., Vol. 2, pp. 1–50.
2. Nagel, R. L. & Gibson, Q. H. (1971) *J. Biol. Chem.* **246**, 69–73.
3. Shim, B. S., Lee, T. H. & Kang, Y. S. (1965) *Nature (London)* **207**, 1264–1267.
4. Kurosky, A., Barnett, D. R., Rasco, M. A., Lee, T.-H. & Bowman, B. H. (1974) *Biochem. Genet.* **11**, 279–293.
5. Kurosky, A., Kim, H.-H., Barnett, D. R., Rasco, M., Touchstone, B. & Bowman, B. H. (1975) *Protides Biol. Fluids, Proc. Colloq.* **22**, 597–602.
6. Kurosky, A., Hay, R. E., Kim, H.-H., Touchstone, B., Rasco, M. & Bowman, B. H. (1976) *Biochemistry* **15**, 5326–5336.
7. Kraut, J. (1977) *Annu. Rev. Biochem.* **46**, 331–358.
8. Stroud, R. M., Kossiakoff, A. A. & Chambers, J. L. (1977) *Annu. Rev. Biophys. Bioeng.* **6**, 177–193.
9. Browne, W. J., North, A. C. T., Phillips, D. C., Brew, K., Vanaman, T. C. & Hill, R. L. (1969) *J. Mol. Biol.* **42**, 65–86.
10. McLachlan, A. D. & Shotton, D. M. (1971) *Nature (London) New Biol.* **229**, 202–205.
11. Delbaere, L. T. J., Brayer, G. D. & James, M. N. G. (1979) *Nature (London)* **279**, 165–168.
12. Shotton, D. M. & Watson, H. C. (1970) *Nature (London)* **225**, 811–816.
13. Birktoft, J. J. & Blow, D. M. (1972) *J. Mol. Biol.* **68**, 187–240.
14. Tulinsky, A., Mani, N. V., Morimoto, C. N. & Vandlen, R. L. (1973) *Acta Crystallogr., Sect. B* **29**, 1309–1322.
15. Huber, R., Kukla, D., Bode, W., Schwager, P., Bartel, K., Diesenhofer, J. & Steigemann, W. (1974) *J. Mol. Biol.* **89**, 73–101.
16. Stroud, R. M., Kay, L. M. & Dickerson, R. E. (1974) *J. Mol. Biol.* **83**, 185–208.
17. Feldmann, R. J. (1976) *Atlas of Macromolecular Structure on Microfiche* (Tracor Jitco, Rockville, MD).
18. Kurosky, A., Barnett, D. R., Lee, T.-H., Touchstone, B., Hay, R. E., Arnott, M. S., Bowman, B. H. & Fitch, W. M. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3388–3392.
19. Jukes, T. H. & Cantor, C. R. (1969) in *Mammalian Protein Metabolism*, ed. Munro, H. N. (Academic, New York), pp. 22–132.
20. Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, DC), Vol. 5.
21. DeHaen, C., Neurath, H. & Teller, D. C. (1975) *J. Mol. Biol.* **92**, 225–259.
22. Katz, L. & Levinthal, C. (1972) *Annu. Rev. Biophys. Bioeng.* **1**, 465–504.
23. Honig, B., Ray, A. & Levinthal, C. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 1974–1978.
24. Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55**, 379–400.
25. Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng.* **6**, 151–176.
26. Malchy, B., Rorstad, O. & Dixon, C. H. (1973) *Can. J. Biochem.* **51**, 265–273.
27. Greer, J. & Bush, B. L. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 303–307.
28. Delbaere, L. T. J., Brayer, G. D. & James, M. N. G. (1979) *Can. J. Biochem.* **57**, 135–144.
29. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112**, 535–542.
30. Chothia, C., Wodak, S. & Janin, J. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 3793–3797.
31. Bloomer, A. C., Champness, J. V., Bricogne, G., Staden, R. & Klug, A. (1978) *Nature (London)* **276**, 362–368.
32. Hwang, P. K. & Greer, J. (1979) *J. Biol. Chem.* **254**, 2265–2270.
33. Rogard, M. & Waks, M. (1977) *Eur. J. Biochem.* **77**, 367–373.
34. Perutz, M. F. (1970) *Nature (London)* **228**, 726–739.
35. Ladner, R. C., Heidner, E. J. & Perutz, M. F. (1977) *J. Mol. Biol.* **114**, 385–414.