We are grateful to Mrs. E. Broich for technical assistance and to Mrs. G. Zubay for assistance in the preparation of the manuscript.

* Research carried out at Brookhaven National Laboratory under the auspices of the U.S. Atomic Energy Commission.

**CHYMOTRYPSINOGEN: INCREASED RESOLUTION AND ABSOLUTE CONFIGURATION***

**By J. Kraut, D. F. High, and L. C. Sieker**

**Department of Chemistry, University of California (La Jolla)**

*Communicated by Bruno H. Zimm, February 12, 1964*

In a previous communication¹ we have reported the early stages of an X-ray diffraction investigation of the structure of type F crystals of bovine chymotrypsinogen A by the method of multiple isomorphous replacement. The three-dimensional Fourier maps described there were based upon intensity data out to a minimum Bragg spacing of 5 Å for the parent crystal and six derivatives. Although information of biochemical significance was disappointingly meager, it was possible at least to observe the outlines of the molecule and to conclude that, in contrast to myoglobin² and hemoglobin,³ few if any straight rods of α-helix are present in chymotrypsinogen. Since it now appears that it will still be some time before a complete detailed structure is forthcoming, we wish to present in this communication something in the nature of a progress report describing our recent efforts to extend the resolution of the Fourier maps from 5 to 4 Å and to establish the absolute configuration of the chymotrypsinogen molecule with the aid of anomalous scattering data.

Although stated thus in terms of minimum Bragg spacing this improvement in resolution would seem very modest, two considerations led us to stop at the 4 Å
stage. First, the number of data to be collected increases approximately as the inverse cube of the minimum Bragg spacing, and therefore it was necessary to measure as many more intensities to reach 4 Å as had already been measured at 5 Å resolution, a task requiring several man-months. Second, and more important, it is evident from the 4 Å data that the phase-determining power of the heavy atoms in our isomorphous derivatives falls off rapidly with increasing resolution, so that little further improvement in the Fourier syntheses might be achieved in spite of the inclusion of many more terms of higher order than 4 Å. As will be seen below, although the improvement has not been dramatic, the 4 Å Fourier maps do appear to be more interpretable than their predecessors.

Data Collection and Phase Refinement.—There are 2383 independent reflections with Bragg spacings greater than or equal to 4 Å for type F crystals of chymotrypsinogen, and thus 1121 additional intensities had to be measured for the native crystal and for each of six heavy-atom derivatives, or a total of 7847. In addition, intensities out to 5 Å were measured for a new HgBr$_4^{2-}$ derivative which showed a somewhat different intensity distribution from the HgI$_4^{2-}$ derivative already in use. Finally, to improve their reliability all 2383 parent intensities were remeasured with new but presumably identical crystals, and the results combined with the previous data.

Integrated intensities were obtained with a G.E. diffractometer as previously described, with the customary precautions against errors due to crystal movement and deterioration. Nine standard reflections were monitored periodically, and their mean intensities used to scale together data from different crystals of the same derivative. Duplicate intensity measurements were made, as indicated above, only for the parent crystals. The mean relative difference in $F$ between these duplicates was 8 per cent; the mean absolute difference on our present approximate scale was 20 e. A mean relative difference in duplicate $F$'s of about 3 per cent would have been expected from counting statistics alone, the larger part of the error probably being due to absorption effects.

A curious feature of the over-all intensity distribution is the presence of a very pronounced dip at sin $\theta/\lambda = 0.080$ (Bragg spacing 6.25 Å) when the local-average intensity is plotted against sin $\theta/\lambda$. This is probably characteristic of crystal structures, like proteins, in which the atoms are "clumped" together and surrounded by rather large interstices containing an effectively uniform electron density. The important point here is that Wilson statistics would probably be inapplicable in such cases, resting as it does on the assumption of randomly and uniformly distributed scattering centers, and this should be borne in mind when attempting to apply statistical arguments to such problems as establishing an absolute scale or determining an over-all temperature factor from low-order data.

A new set of phases based upon the previously determined heavy-atom parameters and a new set of difference-Fourier maps were then calculated, including one for the new HgBr$_4^{2-}$ derivative. The new maps showed a single heavy-atom site for the HgBr$_4^{2-}$, six possible further minor sites for the other derivatives, and reduction in the occupancy of the smallest minor site of the PtCl$_6^{2-}$ derivative to a subsignificant level. Appropriate additions and deletions were made in the list of heavy-atom parameters which was now taken as the starting point for the 4 Å refinement.
Least-squares refinement calculations were carried out on a CDC-1604 with a program very similar to the one previously described. In this series of calculations, however, an individual isotropic temperature factor was included as an adjustable parameter for each heavy-atom site, as well as positional parameters, site occupancies, and individual scale factors for each derivative. Seven cycles of refinement were run, each requiring about 50 min. Six of the smallest minor sites were eliminated when their occupancies became insignificant. All but one of these had been newly introduced just prior to the 4 Å refinement.

The maximum total shift in positional parameter was 1.5 Å; most shifts were about one tenth this magnitude. Some very appreciable changes in occupancies and isotropic temperature factors occurred, however, as, for example, from occupancy 80 to 69 e and from $B = 130$ to 31 Å² for the PtCl$_6^{2-}$ site in the mixed PtCl$_6^{2-}$ and Hgl$_4^{2-}$ derivative. Over-all statistics from the final refinement cycle are as follows:

$$\frac{\sum |kF_H(\text{obs}) - F_H(\text{calc})|}{\sum kF_H(\text{obs})} = 9.5\%$$

with a range of 7–12 per cent for the seven derivatives considered separately;

\begin{align*}
\text{mean } |\varphi \text{ (centroid)} - \varphi \text{ (most probable)}| &= 17.7^\circ \\
\text{r.m.s. } |kF_H(\text{obs}) - F_H(\text{calc})| &= 40 \text{ electrons}
\end{align*}

with a range of 27–51 electrons;

$$\text{r.m.s. } |f_H| = 66 \text{ electrons}$$

with a range of 49–105 electrons; and mean figure of merit, $m = 0.77$.

In the course of refinement the r.m.s. $|kF_H(\text{obs}) - F_H(\text{calc})|$ was reduced only slightly from an initial value of 43 to 40 e, and the mean figure of merit, $m$, increased from 0.72 to 0.77. This improvement in the mean $m$ was considered to be a significant indication that the refinement procedure was working properly since it is only the residual $\sum (kF_H(\text{obs}) - F_H(\text{calc}))^2$ which is directly minimized, the figure of merit being obtained essentially as a by-product of the calculations. As would be expected from the fact that the starting parameters were already refined with respect to the 5 Å data, the over-all statistics for the combined data show little change from what they were for the 5 Å data alone.

When the local average figure of merit for a shell of reciprocal space is plotted as a function of $\sin \theta/\lambda$, it is found to start very close to unity for low $\sin \theta/\lambda$ and to drop off steadily as $\sin \theta/\lambda$ increases, falling to 0.66 at $\sin \theta/\lambda = 0.12$. Such behavior would be expected if the heavy-atom derivatives were not perfectly isomorphous with the parent protein or if the heavy-atom sites were to have effective temperature factors much larger than the over-all temperature factor of the parent protein. The former is probably true in general, and the latter is certainly true in the present case. In fact the calculated effective temperature factors for the various heavy-atom sites range from 31 to 273 Å² with a mean of 150 Å² for 17 sites. These temperature factors were calculated assuming point scatterers for the heavy atoms, and would therefore be about 10 Å² smaller if accurate atomic scattering factors had been used. They are nevertheless much larger than the estimated value of 21 Å² for the temperature factor of the parent protein itself.
Before concluding this section it should also be mentioned that our over-all mean \( m \) of 0.77 is close to the value reported by Cullis et al.\(^7\) for hemoglobin in spite of the fact that the hemoglobin data extended only to 5.5 Å minimum Bragg spacing.

**The 4 Å Fourier Syntheses.**—Three-dimensional Fourier syntheses of the native protein structure were calculated as before with the refined 4 Å phases. Once again the "most probable" and "centroid" Fouriers were in close agreement, and assignment of regions of electron density to individual molecules was reasonably straightforward. Portions of the Fourier maps containing a single molecule were contoured at levels arbitrarily chosen to depict the structure best. Approximately 18 per cent of the unit-cell volume was then found to lie within the lowest contour. The contours were retraced on sheets of transparent film, and these were then taped to Plexiglas plates to provide a convenient three-dimensional display. Figures 1a and 1b are pairs of stereoscopic photographs of portions of a few sections from these Fourier maps. Figure 1a is from the "most probable" Fourier sectioned perpendicular to the \( z \)-axis, and Figure 1b is from the "centroid" Fourier sectioned perpendicular to the \( y \) axis.

It was evident that a significant improvement had resulted from inclusion of the 4 Å data. In many areas we could now see stretches of curved backbone chain with protruding bulges of the size and spacing along the backbone to be expected for side chains. Figures 1a and 1b show this type of feature quite clearly when viewed stereoscopically. It is almost unnecessary to say, however, that in other areas the picture is still confused. The disulfide bridges are still not obvious, and it is still impossible to find a unique path for the backbone chain. There is no evidence of the characteristic straight rods of \( \alpha \)-helix, and in fact even the one short segment of the molecule which had looked like possible \( \alpha \)-helix at lower resolution now appears to be curved and quite probably in a more nearly extended-chain conformation. This particular chain segment is the one shown in Figure 1a. It would be difficult to understand why any \( \alpha \)-helical rods, if they were present, should not show up as clearly as they did in myoglobin and hemoglobin. It is, of course, possible that the backbone chain may occasionally assume an \( \alpha \)-helix-like conformation for a few residues, but it seems doubtful that these regions can extend for more than a single turn of the helix in any one place. In any case, it should be emphasized that until the structure is completely solved and refined any conclusions concerning helix content must remain tentative.

**Absolute Configuration.**—It was by now becoming painfully obvious that in order to make further progress toward obtaining detailed structural information it would probably be necessary to resort to laborious model-building and Fourier refinement procedures. In guessing at a structure for any section of polypeptide chain, however, one is at the outset faced with a fourfold ambiguity: the chain may be running in either direction, and it may be in either the D or the L configuration. We therefore considered establishing the absolute configuration of the molecule as seen in our Fourier maps as an essential preliminary step.

The absolute configuration of myoglobin was determined by taking account of the known absolute configurations of the L-amino acids whose individual side-chains could be resolved in the 2 Å Fourier synthesis.\(^8\) For hemoglobin, on the other hand, where the resolution was inadequate for this approach, the absolute configuration was determined from measurements of anomalous scattering effects due to the iron...
atom of the heme. The two methods proved mutually confirmatory when the conformation of the subunits of hemoglobin turned out to be almost identical with myoglobin.

Although chymotrypsinogen itself contains no atoms with large anomalous scattering components, the heavy-atom derivatives of course do. Indeed, for Cu-Kα radiation $\Delta f''$ is 7.2 e for iodine and 9 e for mercury, while it is only 3.4 e for iron. Furthermore, Rossmann has shown that mercury atoms in substituted hemoglobin may be independently located by Patterson-function methods which take advantage of mercury's anomalous scattering component. It therefore seemed reasonable to expect that the HgI₄⁻⁻ derivative of chymotrypsinogen would show sufficiently large anomalous scattering effects to settle its absolute configuration unambiguously.

---

**Fig. 1.** (a) Stereoscopic view of region in "most probable" Fourier showing example of backbone chain segment. (b) Another example of backbone chain taken from the "centroid" Fourier.
In space group $\text{P}2_1\text{2}1\text{2}_1$ the point symmetry of reciprocal space is mmm if Friedel's law is obeyed and 222 if not. Thus, reflections fall into sets of eight, four related by 222 symmetry with indexes $hkl$, $hk\bar{l}$, $\bar{h}k\bar{l}$, and $\bar{h}\bar{k}\bar{l}$, and four Friedel conjugates of these. Copper $K_\alpha$ intensity data were collected for 78 such sets of eight reflections with the largest expected Friedel differences as calculated from the known HgI$_4^{2-}$ sites, and with $2\theta \leq 9.5^\circ$. The limitation on $2\theta$ was necessary because the G.E. diffractometer does not allow negative $2\theta$'s beyond $-10^\circ$. A single crystal mounting was used. Intensities for each subset of four reflections were averaged to give the basic $F^2(h)$ and $F^2(-h)$ data. The mean relative Friedel difference was 14 per cent.

The simplest way to determine absolute configuration from such data is to compare the observed signs of the Friedel differences $F^2(h) - F^2(-h)$ with those calculated on the basis of the original arbitrarily assumed enantiomorph. In the present case 69 out of the 78 observed differences were of opposite sign, indicating the correct configuration of the chymotrypsinogen molecule to be the mirror image of the balsa-wood model shown in Figures 1 and 2 of our previous publication.

This conclusion is confirmed by a more elaborate treatment of the data wherein a 78 term "anomalous difference Fourier" synthesis was calculated and found to have a prominent negative peak at the previously assigned HgI$_4^{2-}$ site. The anomalous difference Fourier function is given by

$$\Delta \rho'' = \frac{1}{V} \sum [F_o(h) - F_o(-h)] \sin [\varphi(h) - 2\pi h \cdot x].$$

Conclusions.—The obvious task facing us now is to improve the Fourier maps still further. Intriguing though occasional features like those of Figures 1a and 1b are, they are still very far from the kind of detailed picture we must have in order to understand the mechanism of the proenzyme-enzyme system in terms of molecular structure.

Unfortunately, there are several indications, some of which have already been mentioned, that the present heavy-atom derivatives would not be of much use for phase determination beyond 4 Å. For one, it is evident from comparison of precession photographs of the derivatives and the parent protein that intensities of higher-order reflections have not been affected by the heavy atoms. More quantitatively, there is the observed drop-off in local average figure of merit with increasing $\sin \theta/\lambda$. Finally, it should be re-emphasized that even for the reflections already included in the 4 Å Fourier maps the r.m.s. calculated heavy-atom contribution to the derivatives' structure factors is only 66 e, just 3 times the error of measurement as established by replication of the parent data, and 1.5 times the r.m.s. difference between observed and calculated reflection amplitudes for the derivatives. These findings are, of course, correlated with the very large apparent temperature factors calculated for the heavy-atom sites.

It is likely that one would obtain a significant improvement in this respect if heavy-atom derivatives could be made which consisted of purified individual protein compounds with covalently bound heavy atoms at single well-defined sites. Then too, it would probably be very helpful if accurate absorption corrections could be made for every measured intensity. Absorption effects are in all likelihood the most important source of experimental error.
In the present case, however, we are inclined to give first priority to another possible approach: model building and classical Fourier refinement. The general impression one gets upon studying the 4 Å Fourier maps is that it should be possible with their help to arrive at detailed and approximately correct structure for portions of the molecule if the known stereochemical requirements of polypeptide chains are carefully followed. Then with this part of the structure as input one ought to be able to follow the time-honored crystallographic procedure of Fourier refinement to correct that input and to reveal further structure. Whether enough of the structure can be correctly deduced initially from the present Fourier maps to permit the refinement to "lock in," and whether it is feasible to locate and record the coordinates of more than 1700 atoms several times over in a reasonable length of time are questions which can only be answered by making the attempt. This we are now doing.

**Summary.**—A new set of Fourier maps have been calculated for type F chymotrypsinogen from phases obtained by multiple isomorphous replacement, including terms out to a minimum Bragg spacing of 4 Å. Features can now be seen which seem to be individual curved backbone chains with side-chain bulges. There is nothing which would correspond to an α-helix segment of more than a single turn. Disulfide bridges are still not obvious, and it is still not possible to follow a unique path for the backbone chain.

The absolute configuration of the chymotrypsinogen molecule has been established with the aid of anomalous scattering data from the HgI$_4^{--}$ derivative. The correct configuration is the mirror image of our previously published model.

It appears unlikely that our present heavy-atom derivatives would enable the resolution to be extended much further. Possible approaches to the complete structure are discussed.

*This project has been supported by research grants GM-10928 from the U.S. Public Health Service and GB-758 from the National Science Foundation, and by grants of computer time from the Research Committee of the University of California, San Diego. We are grateful to Dr. Herman Cohen of the Princeton Laboratories, Inc., for continuing to supply us with chymotrypsinogen gratis.


12 Kraut, J., to be published.