The σ54 RNA polymerase holoenzyme functions in enhancer-dependent transcription. The structural organization of the σ54 subunit of bacterial RNA polymerase in solution is analyzed by synchrotron x-ray scattering. Scattering patterns are collected from the full-length protein and from a large fragment able to bind the core RNA polymerase, and their low resolution shapes are restored using two ab initio shape determination techniques. The σ54 subunit is a highly elongated particle, and the core binding fragment can be unambiguously positioned inside the full-length protein. The boomerang-like shape of the core binding fragment is similar to that of the atomic model of a fragment of the Escherichia coli σ70 protein, indicating that, although the σ54 and σ70 factors are unrelated by primary sequence, they may share some structural similarity. Potential DNA binding surfaces of σ54 are also predicted by comparison with the σ54 core binding fragment.

Transcriptional regulation of gene expression is a major area where information flow from DNA is controlled, often by activation of RNA polymerase function, and is central to regulating the initiation rates (1, 2). The expression of diverse gene sets in bacteria requires the specialized σ54-containing RNA polymerase (2), likened in its properties to the eukaryotic RNA polymerase II (3). Transcription activation involves a poorly defined interaction of activator proteins, bound at remote sites, with the σ54 holoenzyme bound at a promoter as a transcriptionally inactive closed complex (4, 5). Our objectives are to determine the contributions of the σ54 protein to RNA polymerase function and activation through structure-function studies. The σ54 protein has a modular domain organization (4–8), and individual domains can be prepared as active partial σ54 sequences, potentially allowing the assignment of structural features seen in the complete σ54 molecule (9).

σ factors contribute to several functions of the holoenzyme, and although discrete activities are known to reside within individual domains, the domains appear to interact with each other for the full function of the σ54. In the present paper, the structural organization of the 54-kDa monomeric σ54 protein from Klebsiella pneumoniae bacteria is analyzed using x-ray solution scattering (SAXS), the method yielding low resolution structural information at nearly physiological conditions. Two methods of ab initio shape restoration (10–12) are used to establish the shapes of the full-length protein and of its 30-kDa fragment (previously obtained as a stable product of σ54 proteolysis). The latter shape can be unambiguously positioned within the former, providing insight about the structural organization of the σ54 molecule. Further, the shape of the 30-kDa fragment exhibits similarity with that of the crystal structure of a fragment of the Escherichia coli σ70 protein that also binds the core RNA polymerase. A prediction of potential σ54 DNA binding surfaces is also made.

EXPERIMENTAL PROCEDURES

Sample Preparation—K. pneumoniae σ54 protein (amino acids 1–477) and the 30-kDa core binding fragment (amino acids 70–324) with an 11-amino acid amino acid extension (MARIRARQSGSR) at its N terminus were overproduced as soluble proteins in E. coli and purified as described previously (13). They were concentrated to 15–20 mg/ml and buffer-exchanged into 10 mM Tris-HCl, 5% glycerol, pH 8.0, at 4 °C by centrifugal ultrafiltration. Concentrated samples were stored at −70 °C and thawed on ice prior to the scattering experiments. Protein concentrations of σ54 and the 30-kDa fragment were determined side by side using the Bio-Rad dye assay and bovine serum albumin as standard.

Scattering Experiments and Data Treatment—The synchrotron radiation x-ray scattering data were collected using standard procedures on the X33 camera (14–16) of the EMBL on the storage ring DORIS III of the Deutsches Elektronen Synchrotron (DESY) and multiwire proportional chambers with delay line readout (17). Samples at concentrations between 2 and 15 mg/ml were measured at a wavelength λ = 0.15 nm for sample-detector distances of 2.9 and 1.4 m covering the momentum transfer ranges 0.20 < s < 2.0 nm<sup>−1</sup> and 0.35 < s < 5.0 nm<sup>−1</sup>, respectively (s = 4πsinθ/λ, where 2θ is the scattering angle). The data were normalized to the intensity of the incident beam, corrected for the detector response; the scattering of the buffer was subtracted; and the difference curves were scaled for concentration using the program SADPAK. The curves recorded at a sample-detector distance of 2.9 m were extrapolated to zero concentration and merged with the data obtained at 1.4 m.

The maximum dimensions D<sub>max</sub> of the σ54 and the 30-kDa core were estimated from the experimental data using the orthogonal expansion program ORTOGONOM (19). The distance distribution functions q<sub>r</sub> and the radii of gyration R<sub>g</sub> were evaluated by the indirect Fourier transform program GNON (20–21). The molecular masses of the solutes were estimated by comparison of the extrapolated forward scattering I(0) with that of a reference solution of bovine serum albumin.

Prior to the shape analysis, undesirable contributions from the scattering due to the internal particle structure at higher scattering angles that become significant above approximately s = 2.0 nm<sup>−1</sup> were re-
ing curves in the ranges up to $s_{\text{max}} = 2.8 \text{ nm}^{-1}$ (α54) and 3.3 nm$^{-1}$ (30-kDa core) were used to compute the excluded volumes $V$ of the hydrated particles (22) and for the *ab initio* shape determination. The outer portions of the curves dominated by the scattering from the internal structure were discarded in the further analysis.

**Shape Determination**—The low resolution particle shapes were re-
stored from the experimental data using two *ab initio* procedures. In the
method of Svergun *et al.* (10, 11), the shape is represented by an
angular envelope function $r = F(\omega)$, where $(r, \omega)$ are spherical
coordinates. The envelope is parameterized as follows (23),

$$ F(\omega) = \sum_{l=0}^{L} \sum_{m=-l}^{l} f_{lm}(\omega) Y_{lm}(\omega) $$

(Eq. 1)

where $Y_{lm}(\omega)$ are spherical harmonics and the multipole coefficients $f_{lm}$
are complex numbers. The maximum order of the spherical harmonics
$L$ is selected to keep the number of free parameters $M = (L + 1)^2 - 6$
close to the number of Shannon channels $N_s = D_{\text{max}}^2/\pi$ in the
experimental data (10, 24–25). The scattering intensity of the envelope
is as follows,

$$ I(s) = 2\pi^2 \sum_{i=0}^{L} \sum_{m=-l}^{l} |A_{lm}(s)|^2 $$

(Eq. 2)

where the partial amplitudes $A_{lm}(s)$ are calculated from the coefficients
$f_{lm}$ using the recurrence relation of Svergun and Stuhrmann (26). These
coefficients are determined by the program SASHA (10, 11) by mini-
mizing the discrepancy $\chi$ between the calculated and the experimental
curves as follows,

$$ \chi = \left[ \frac{1}{N} \sum_{j=1}^{N} \left( \frac{I(s_j) - \overline{I}(s_j)}{\sigma(s_j)} \right)^2 \right]^{1/2} $$

(Eq. 3)

where $N$ is the number of the experimental points and $\overline{I}(s_j)$ and $\sigma(s_j)$
are the experimental intensity and its S.D. in the $k$th point, respec-
tively. The spatial resolution of the envelope function representation is
approximately $\delta r = \sqrt{5/\pi R_1}/(\sqrt{3} (L + 1))$.

Another *ab initio* procedure employs a particular implementation of
a general method (12). A sphere of diameter $D_{\text{max}}$ is filled by densely
packed small spheres (dum
ymy atoms) of radius $r_{\text{core}} < D_{\text{max}}$. The struc-
ture of this dummy atoms model (DAM) is defined by a configuration $X$,
assigning an index to each atom corresponding to solvent (0) or solute
particle (1). The scattering intensity from the DAM is computed using
Equation 2 with the following partial amplitudes,

$$ A_{lm}(s) = i^l \sqrt{2\pi} \sum_{j} j(\sigma r) Y_{lm}(\omega) $$

(Eq. 4)

where the sum runs over the dummy atoms with $X_j = 1$ (particle atoms); $r_{\text{core}}, \omega_j$ are their polar coordinates; and $j(\sigma x)$ denotes the spherical Bessel
function. The number of dummy atoms in the search model $M_{\text{DAM}} = (D_{\text{max}}/r_{\text{core}})^3$ is significantly exceeded by the number of Shannon channels.
To reduce the effective number of free parameters in the model, the
method searches for a configuration $X$ minimizing $\beta(X) = \beta^2 + \alpha P(X)$,
where the looseness penalty $P(X)$ with a positive weight $\alpha > 0$ ensures
that the DAM has low resolution with respect to the packing radius $r_{\text{core}}$.
The weight of the penalty is selected to have significant penalty contri-
bution at the end of the minimization. The latter is performed starting
from a random approximation using the simulated annealing method
(27); details of the procedure are described elsewhere (12). The spatial
resolution provided by the DAM is defined by the range of the moment-
um transfer as $\delta r = 2m_{\text{max}}/\pi D_{\text{max}}$.

Since the shape determination methods yield models with an arbi-
trary orientation and handedness, they were appropriately rotated for
comparison with each other and also with the atomic model of a σF0
subunit fragment from *E. coli* RNA. The latter was taken from the
Protein Data Bank (28), entry 1sig (29). The model of the fragment of
the B-DNA was generated with the program Turbo-FRODO (30), and
the three-dimensional models were displayed using the program ASSA
(31).

**RESULTS**

The experimental scattering curves from the α54 and its
30-kDa fragment are presented in Fig. 1, and the structural
parameters computed from these curves are listed in Table I.
The molecular masses of the solutes indicate that both σ54 and the 30-kDa fragment are monomeric in solution, in agreement with the results of analytical ultracentrifugation. The experimental values of $D_{\text{max}}$ and $R_g$ suggest that the two macromolecules, and especially σ54, are rather anisometric. The volumes of the hydrated σ54 and the 30-kDa fragment are twice the dry volumes expected from the molecular masses (about 75 and 42 nm$^3$, respectively). This reveals a very high hydration (0.4 g of H$_2$O/g of protein) characteristic for extended macromolecules with high specific surface accessible to the solvent. The profiles of the distance distribution functions $p(r)$ in Fig. 2 are typical for elongated particles (32). It follows that σ54 is a very elongated protein and that the 30-kDa fragment, although having a 2-nm smaller maximum diameter, is still rather elongated.

The portions of the scattering curves used for *ab initio* shape determination contained $N_s = 11.6$ Shannon channels both for σ54 and for the 30-kDa fragment, allowing the use of harmonics up to $L = 4$ (19 independent parameters) in the shape determination using the envelope functions (10). The restored envelopes are presented in Fig. 3 (*left column*), and the fits to the experimental data are displayed in Fig. 1, *a* and *b*. The envelope of the 30-kDa fragment can be unambiguously positioned inside that of σ54. Moreover, a boomerang-like shape of the core correlates well with the gross structure of the core RNA polymerase-binding fragment of the *E. coli σ70* protein (the atomic model of the latter is superimposed with the low resolution models in Fig. 3). Nevertheless, the shape reconstruction using envelope functions has its limitations for elongated particles like σ54. To keep the model at low resolution and the number of model parameters reasonably small, higher spherical harmonics are omitted in the description of the envelope (Equation 1). This ensures uniqueness of the model restoration (10) but leads to a termination effect especially noticeable for anisometric particles for which the higher harmonics play a significant role. As a result, the radius of gyration and the maximum size of the restored envelopes are lower than those determined experimentally (Table I), and the calculated curves display systematic deviations from the experimental data (Fig. 1).

A much less restricted description of the particle shape is provided by the alternative DAM technique. The search volume for the σ54 contained $M_{\text{DAM}} = 2243$ dummy atoms with a packing radius $r_p = 0.45$ nm within a sphere with the diameter $D_{\text{max}} = 13$ nm. Of those, 295 dummy atoms were attributed to the particle in the final model of σ54, and they further defined the search volume for the 30-kDa fragment. The obtained DAMs presented in Fig. 3 (*right column*), yield better fits to the experimental data in Fig. 1 than those provided by the envelope models and also the structural parameters that neatly agree with the experimental values (Table I). To verify the uniqueness of the shape restoration using DAM, several independent restorations were performed using different conditions (by varying the packing radius $r_p$ and by starting from random initial configurations). These restorations yielded reproducible results similar to that in Fig. 3 (*right column*). In particular, the shape of the 30-kDa fragment restored *ab initio* within the spherical search volume with $D_{\text{max}} = 11$ nm was very similar to that presented in Fig. 3.

### DISCUSSION

An *ab initio* shape restoration from solution scattering data is not unique unless the search model is kept at low resolution (10). The two *ab initio* procedures employed here use different restrictions. In the envelope restoration, higher spherical harmonics are omitted in the description of the envelope function so that the number of the adjustable parameters is comparable with that of the Shannon channels in the scattering data. In the DAM retrieval, the number of parameters is much larger ($M_{\text{DAM}} \sim 10^3$), and the looseness penalty ensures that the final model is compact and interconnected. Given the differences in parameterization and in the numbers of model parameters, similarities in the low resolution models independently restored by the two methods are remarkable. The differences between the two restorations in Fig. 3 are largely attributed to the termination effects in the envelope method, which are

### TABLE I

| Parameter/Sample | σ54 | 30-kDa fragment |
|-----------------|-----|-----------------|
|                 | Exp. | Env. | DAM | Exp. | Env. | DAM |
| Mass (kDa)      | 58 ± 6 |     |     | 32 ± 4 |     |     |
| $D_{\text{max}}$ (nm) | 13.0 ± 0.5 | 10.4 | 12.5 | 11.0 ± 0.5 | 8.3 | 11.0 |
| $R_g$ (nm)      | 3.43 ± 0.06 | 3.04 | 3.39 | 3.05 ± 0.05 | 2.51 | 3.08 |
| $V$ (nm$^3$)    | 155 ± 10 | 144 | 152 | 85 ± 6 | 72.2 | 86.7 |
| Free parameters | 11.6$^b$ | 19 | 2243 | 11.6$^d$ | 19 | 295 |
| Discrepancy $\chi$ | 1.79 | 1.18 |     | 2.8 | 2.2 |     |
| Resolution (nm) |     |     |     | 1.49 | 0.88 |     |

$^a$ Exp, calculated from the experimental data.
$^b$ Env, parameters of the envelope models.
$^c$ DAM, parameters of the dummy atoms models.
$^d$ Number of Shannon channels is given.
significant for elongated particles like $\sigma54$. It should be stressed that although the shapes obtained with the DAM technique are still low resolution models, their better nominal resolution (Table I) and more detailed appearance compared with the envelope models are fully justified by better fits to the experimental data.

DNA footprinting experiments have shown that $\sigma54$ interacts with promoter DNA from $-33$ to $-5$. Major points of contact are within four base pairs (TGCA) centered at $-12.5$ and seven (CTGGCAC) centered at $-24$, within the consensus sequence of $\sigma54$ DNA-binding sites. The distance between these contact centers suggests that unless $\sigma$ binds to a non-B-DNA conformation, interactions between $\sigma$ and DNA will involve contacts made by structures within more than one of the well developed lobes revealed by the SAXS analysis. Superposition of the low resolution dummy atoms models with a 28-base pair B-DNA fragment in Fig. 3 (top right orientation) summarizes a relationship between $\sigma54$, the 30-kDa fragment, and promoter DNA that can account for the known extent of $\sigma$-promoter DNA interactions. The 30-kDa fragment analyzed by SAXS lacks the $\sigma$ DNA-binding domain and region I sequences that also contribute to DNA recognition. It therefore seems probable that the lobe missing from the 30-kDa fragment (Fig. 3) but present in $\sigma54$ includes some of the structures that direct promoter binding. The upper surface of $\sigma54$ depicted in Fig. 3 (top right orientation) further distinguishes the full-length protein from the 30-kDa fragment and may contribute additional DNA contacting structures. Together, the lobe and the upper surface can provide DNA-contacting surfaces of approximately the necessary extent to fully interact with promoter DNA. The direction of the DNA sequence displayed in Fig. 3 is arbitrary, since no experimental data correlating which structure recognizes the $-12$ or $-24$ sequences are available.

The low resolution solution structure for $\sigma54$ determined in this work provides the first initial model for a protein that functions to convert E. coli RNA polymerase into an enhancer-dependent enzyme (3). Clearly developed structures are present that probably represent either a discrete individual domain or alternately domains that are directly interacting within the

![Fig. 3. Low resolution models of $\sigma54$ (yellow) and 30-kDa fragment (red). Left, envelope models (semitransparent envelopes); right, DAMs (semitransparent spheres of radius $r_w = 0.45$ nm). The middle and bottom rows are rotated counterclockwise by 90° around the x and y axes, respectively. The models are superimposed with the crystal structure of the $\sigma70$ subunit fragment from E. coli RNA (blue Cα trace). A 28-base pair-long B-DNA fragment comprising the $\sigma54$ binding site (green) is shown above potential DNA binding surfaces (top orientation in the right panel).](image)
tertiary structure. The amino-terminal domain required for transcription silencing and activator responsiveness and the C-terminal DNA binding domain are thought to interact (9). Comparison of the structure of the 30-kDa fragment that lacks both of these domains with the full-length protein structure is consistent with such an interaction, since one lobe is clearly absent (Fig. 3). Possibly, both the amino-terminal domain and C-terminal domain contribute to this lobe and do directly interact with each other. Another striking observation is that the fragment of σ70 for which the atomic structure has been determined has a similar shape and size compared with the 30-kDa fragment of σ54 (Fig. 3). Both fragments bind to the same core RNA polymerase. Assuming the shape of the fragments is related to core binding, it seems that σ70 and σ54 are probably located in similar places on the core. The sequence differences that exist between σ54 and σ70 are, however, sufficient to account, at least in part, for the different properties of the two holoenzymes. Some interactions with core may nevertheless be conserved between the two σ types.

CONCLUSION

The low resolution solution structure of σ54 shows a well developed anisometric shape. Similarity to a part of σ70 that binds to the core RNA polymerase and interacts with the −10 promoter DNA is evident. Now that crystallography and cryo-electron microscopy are beginning to reveal the structures of multisubunit RNA polymerases (18), one can also begin to evaluate the functional significance of this similarity. These structural approaches together with the use of tethered iron chelate methods for examining proximity relationships within transcription complexes should help determine the organization of RNA polymerase holoenzymes.

REFERENCES

1. Helmann, J. D., and de Haseth, P. L. (1999) Biochemistry 38, 5959–5967
2. Merrick, M. J. (1993) Mol. Microbiol. 10, 903–909