

# Reports

## Multiple Conformational States of Proteins: A Molecular Dynamics Analysis of Myoglobin

R. ELBER AND M. KARPLUS

A molecular dynamics simulation of myoglobin provides the first direct demonstration that the potential energy surface of a protein is characterized by a large number of thermally accessible minima in the neighborhood of the native structure (for example, approximately 2000 minima were sampled in a 300-picosecond trajectory). This is expected to have important consequences for the interpretation of the activity of transport proteins and enzymes. Different minima correspond to changes in the relative orientation of the helices coupled with side-chain rearrangements that preserve the close packing of the protein interior. The conformational space sampled by the simulation is similar to that found in the evolutionary development of the globins. Glasslike behavior is expected at low temperatures. The minima obtained from the trajectory do not satisfy certain criteria for ultrametricity.

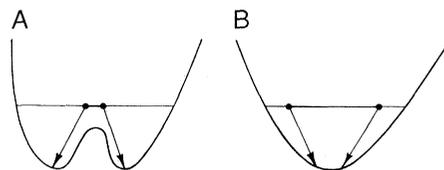
THE INTERNAL MOTIONS OF PROTEINS are being studied both for their intrinsic interest (proteins are disordered systems with special characteristics) and their functional role (1). For native proteins with a well-defined average structure, two extreme models for the internal motions have been considered. In one, the fluctuations are assumed to occur within a single multidimensional well that is harmonic or quasiharmonic as a limiting case (2–5). The other model is based on the assumption that there are multiple minima or substates; the internal motions correspond to a superposition of oscillations within the wells and transitions among them (3, 6–10). Experimental data have been interpreted with both models, but it has proved difficult to distinguish between them (11).

In this report, we demonstrate that multiple minima exist in proteins and show that they have an essential role in determining their internal motions. A molecular dynamics simulation of the protein myoglobin is used to characterize the minima structurally and energetically. Myoglobin was chosen for study because it has been examined experimentally by a variety of methods and the two motional models have been applied to it (5–8, 12, 13). It is ideally suited for analysis, because its well-defined secondary structure (a series of  $\alpha$ -helices connected by loops) facilitates a detailed characterization of the dynamics. The widespread interest in the multiple minimum problem in amorphous and glassy materials makes this study timely (14–19), particularly because it has been suggested that proteins have glasslike properties (14–18).

We concentrate on the structural and dynamic aspects that are of primary interest

for inhomogeneous systems like proteins. The results provide a sound basis for motional models of proteins and make possible an evaluation of their glasslike nature and possible ultrametric character. Further, the range of structural fluctuations found in the dynamics of a single myoglobin molecule is of the same order of magnitude as differences among the structures for a series of globins (20).

The analysis is based on a 300-psec classical molecular dynamics simulation of myoglobin at 300 K; details of the simulation method have been presented (21). The topography of the potential surface underlying the dynamics was explored by finding the local energy minima associated with sequential coordinate sets (16, 17). Thirty-one coordinate sets (one every 10 psec) were selected, and the energy of each was minimized with a modified Newton-Raphson algorithm suitable for large molecules (22); 2000 or more steps were required to obtain a minimized structure with an atomic potential energy gradient of 0.06 kcal/(mole  $\text{\AA}$ ). Since the coordinate sets all corresponded to different minima, structures separated by



**Fig. 1.** Representation of the rms difference criterion for different minima. **(A)** The rms after the minimization is larger than the initial rms, implying that the two conformations correspond to different minima; **(B)** rms after the minimization is smaller than the initial rms, implying that the two conformations correspond to the same minimum.

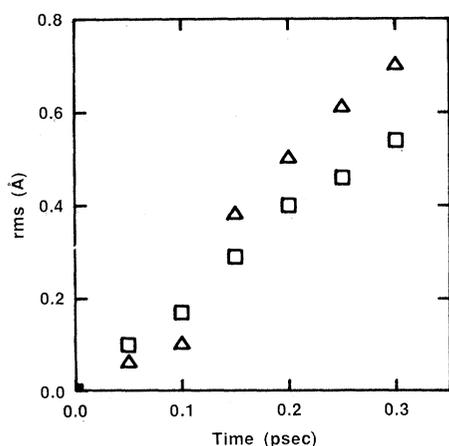
shorter time periods were examined to determine how long the trajectory remains in a given minimum. Seven additional coordinate sets (one every 0.05 psec) were chosen, and their behavior on minimization was examined; if two coordinate sets converged, they corresponded to the same minimum; if they diverged, they corresponded to different minima (Fig. 1). The measure for the distance between two structures is their root-mean-square (rms) coordinate difference after superposition, where

$$\text{rms} = [(1/N) \sum_{i=1}^N \{(x_i)_a - (x_i)_b\}^2]^{1/2} \quad (1)$$

with  $(x_i)_a$  and  $(x_i)_b$  the coordinates of atom  $i$  in structures  $a$  and  $b$ , respectively; the sum is over the  $N$  atoms in the molecules (excluding the heme group). The change in rms difference between a pair of structures on minimization was in the range 0.06 to 0.7  $\text{\AA}$ , relative to the initial difference of 0.1 to 0.5  $\text{\AA}$  for the structures separated in time by 0.05 to 0.35 psec.

Analysis of the short time dynamics (Fig. 2) demonstrates that convergence occurs for intervals up to  $0.15 \pm 0.05$  psec, the uncertainty corresponding to the time between coordinate sets. Thus, the 300-psec simulation samples about 2000 different minima; this is a sizable number, but it may nevertheless be small relative to the total (finite) number of minima available to such a complex system in the neighborhood of the native average structure (that is, conformations that are natively like and significantly populated at room temperature); whether this number increases exponentially with the number of particles, as has been shown for spins (18) and other simple many-body systems (16) is not known for proteins. The rms differences among the minimized structures reach a maximum value of approximately 2  $\text{\AA}$  at about 200 psec. Thus, the vector  $\mathbf{R}_K$  representing the coordinates of all the atoms in a natively like conformation  $K$  is restricted to a volume such that the distance between any two minimized conformations  $\mathbf{R}_K$  and  $\mathbf{R}_{K'}$  is limited.

Comparison of the energies of the minimized structures (Fig. 3) shows that width of the energy distribution is on the order of 20 K (40 cal/mol) per degree of freedom. Since the difference in energy between the "inherent" structures (16, 17) is small, they are significantly populated at room temperature. Further, the large number of such structures sampled by the room temperature simulation suggests that the effective barriers separating them are low and that the protein is undergoing frequent transitions



**Fig. 2.** Root-mean-square difference correlation function  $\langle \text{rms}(0), \text{rms}(t) \rangle$  between structures as a function of the time interval between them. Results are for 0.05-psec time interval structures; the squares are from the trajectory and the triangles are the minimized structure results.

from one structure to another. The fluctuations within a well can be described by a harmonic or quasiharmonic model. Estimates based on the time development of the rms atomic fluctuations for main-chain atoms at room temperature (23) indicate that 20 to 30 percent of the rms fluctuations are contributed by oscillations within a well and 70 to 80 percent arise from transitions among wells; for side chains the contribution from transitions among the multiple wells is expected to be larger. These results are similar to those obtained in a simulation of an isolated helix (8).

From the results, extrapolated to lower temperatures in terms of a temperature-independent potential, the expectation is that the oscillations within a well would decrease linearly or nearly linearly with temperature [until the quantum regime is reached (8)], while the transitions among wells would go to zero [except for possible tunneling contributions (24)] when the barriers

become so high relative to the thermal energy that individual molecules are trapped in separate wells. Activation energies in the range of 2 to 8 kcal/mol have been suggested (14). Since energy differences among some of the wells are small (Fig. 3), molecules may be trapped in metastable states at low temperatures, in analogy to violations of the third law of thermodynamics in crystals (such as crystals of CO) and models for the glassy state (14–19). A number of experiments suggest that the transition temperature for myoglobin is in the neighborhood of 200 K (6, 9, 14, 25). Because large-scale, collective motions that involve the protein surface are important in the fluctuations (23), it is likely that the observed transition is due to the freezing of the solvent matrix (23, 25).

Since the details of the native structure of a protein play an essential role in its function, it is important to determine the structural origins of the multimimum surface obtained from the dynamics analysis. The superposed stereopictures of two states separated by 100 psec (Fig. 4) illustrate that the general features of the structure (helices and turns) are preserved throughout the simulation and that the differences in position are widely distributed. To analyze the structural changes, a distance matrix approach was used (26); the elements of this matrix are  $R_{ij}(K)$ , the distance between particles  $i$  and  $j$  in conformation  $(K)$ . Of interest for comparing two conformations  $K$  and  $K'$  is the difference distance matrix

$$\Delta_{ij}(K, K') = R_{ij}(K) - R_{ij}(K') \quad (2)$$

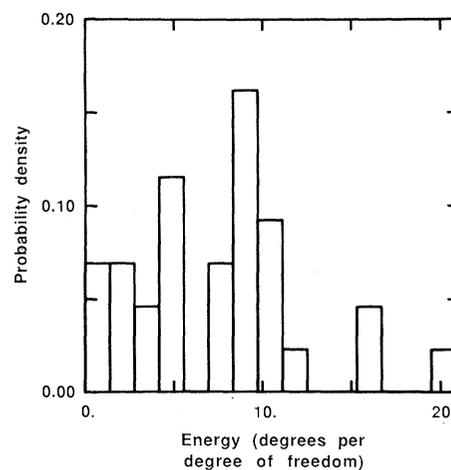
The scaled difference matrix  $\{\Delta_{ij}\}$ , is the dimensionless matrix obtained by dividing  $\Delta_{ij}(K, K')$  by the largest element. Figure 5 for the  $C\alpha$  atoms shows that the points corresponding to the large  $\{\Delta_{ij}\}$  values tend to occur in groups on straight lines parallel to the axes. They are associated primarily with loop displacements or the relative motions of helices which individually behave as nearly rigid bodies. For times of 100 psec and greater the difference matrices are similar in that the dominant features are present in all of the matrices; below 100 psec, changes occur in one or more loop regions

**Table 2.** The relative motions of helices involved in contacts. The fluctuations in the distances (in angstroms) and angles (in degrees) between pairs of helices are given. Listed are the rms fluctuations, the maximum differences found in the set of minimized conformations, and the differences between structures in the globin series.

Helix contact	Trajectory				X-ray structures*	
	$\langle \Delta r^2 \rangle^{1/2}$	$\langle \Delta \theta^2 \rangle^{1/2}$	$\Delta r_{\max}$	$\Delta \theta_{\max}$	$\Delta r$	$\Delta \theta$
A-H	0.7	2	2.2	10	2.0	12
B-E	0.4	14	1.6	39	3.0	25
B-G	0.4	6	1.9	25	6.2†	30
F-H	0.4	1	1.8	5	1.4†	10†
G-H	0.3	2	1.3	7	2.5	15

\*From (20).

†Calculated as described in text.



**Fig. 3.** Energy distribution of the minimized conformations; the energy is in degrees Kelvin per degree of freedom.

but not all of the loop or helix relative motions appear in the difference matrix between any pair of structures. This suggests that rearrangements within individual loops are the elementary step in the transition from one minimum to another and that these are coupled with associated helix displacements. Which loop or turn changes in a given time interval appears to be random. Specific loop motions may be initiated by side-chain transitions in the helix contacts, main-chain dihedral angle transitions of the loops themselves, or a combination of the two. As the time interval between two structures increases, more loop transitions have occurred. At room temperature, the transition probabilities are such that for an interval 100 psec or longer between two structures, some transitions will have taken place in all of the flexible loop regions and the scaled difference matrix representation has converged. However, since the rms differences between structures continue to increase up to 200 psec, the configuration space available to the molecule includes a range of structures for the loop regions that are not completely sampled in 100 psec.

To characterize the helix motions that are coupled with the loop rearrangements, the internal structural changes of the helices were separated from their relative motions.

**Table 1.** Internal and relative fluctuations of helices. The procedures are described in the text.

Regions	Internal rms (Å)*	Translation (Å)	Rotation (deg)
Nonhelical regions†	2.5 ± 0.4	1.4	6
Helices			
A	0.9 ± 0.3	1.8	9
B	0.9 ± 0.4	2.1	17
C	0.8 ± 0.3	2.0	18
D	0.4 ± 0.1	5.4	17
E	1.1 ± 0.3	2.2	11
F	0.5 ± 0.2	2.2	21
G	0.7 ± 0.2	4.8	17
H	0.8 ± 0.4	1.3	11

\*The rms values and their standard deviations are given. †The results for the nonhelical regions are averages over the loops and turns of the molecule.

Individual helices and loops were superimposed, and the rms differences for the main chain were calculated for the set of structures; the rms difference for the internal structure of the helices is generally less than 1 Å (see Table 1). Helix E undergoes somewhat greater internal structural changes; this appears to be due to significant bending of the helix during the trajectory. The corre-

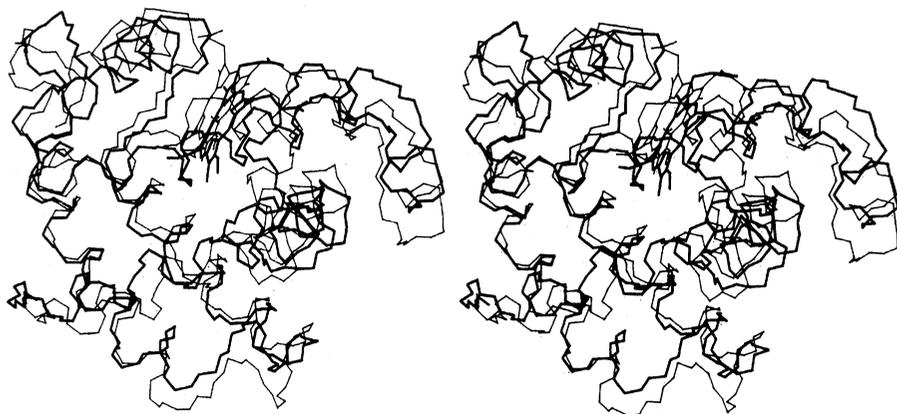
sponding results for the loop regions show that they undergo much larger internal structural changes. If the helices are superimposed successively according to their sequential order in the chain (that is, first helix A is superposed in two structures and the rotation and transition of the molecules to superpose helix B is determined; having superposed helix B, the rotation and transla-

tion required to superpose helix C is determined, and so on), the required rms translations fall into two ranges (see Table 1); they are of the order of 1.5 to 2 Å, except for C compared with D and F with G, which are of the order of 5 Å; the corresponding angular variations are between 11° and 21° and show no simple correlation with the translations. The behavior of the C and D helices can be correlated with the flexibility of the C-D loop in myoglobin; crystal structures show alternative conformations for this region (27).

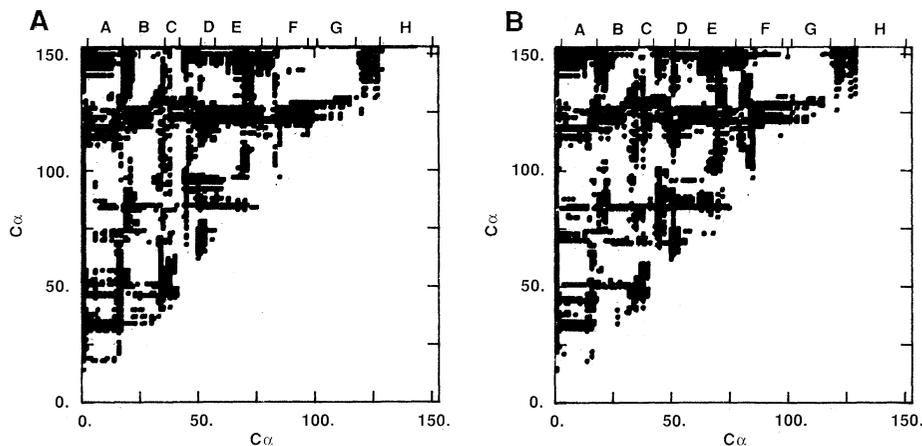
Pairs of helices that have significant van der Waals interaction were also examined (Fig. 6) (20); these are helix pairs A-H, B-E, B-G, F-H, and G-H, for all of which at least three residues from each helix are in contact. Each helix is fitted to a straight line and the fluctuations of the distance between the helix centers of mass and the relative orientations of the lines are compared (Table 2). The relative translations found in this case have rms values of 0.3 to 0.7 Å, and the relative rotations have rms values of 1° to 14°. The maximum differences between pairs of structures are 1.3 to 2.2 Å and 5° to 39°. These values are of the same order as the differences (1 to 3 Å, 15° to 30°, except for the B-E contact where the maximum difference is 6 Å) found in comparing a series of different globins with known crystal structures and sequence homologies in the range 16 to 88 percent (20). Thus, the range of conformations sampled by a single myoglobin trajectory is similar to that found in the evolutionary variation among crystal structures of the globin series.

The comparison of the various globin structures (20) suggested that the range of different stable helix packings is achieved primarily by changes in side-chain volumes resulting from amino acid substitutions. In the dynamics, it is the correlated motions of side chains that are in contact, plus the rearrangements of loops, that make possible the observed helix fluctuations. Different positions within wells and transitions between wells for side chains (for example,  $\pm 60^\circ$ ,  $180^\circ$  for  $\chi_1$ ) are involved. This is in accord with the results of high-resolution x-ray studies that show significant disorder in side-chain orientations (27, 28). Further, correlated dihedral angle changes differentiate the various minima; for example, for the B-E helix contact, residues Glu<sup>26</sup> (B), Leu<sup>61</sup> (E) and Val<sup>21</sup> (B), Leu<sup>69</sup> (E) are involved in correlated motions. Since more than one set of side-chain orientations is consistent with a given set of helix positions, the known globin crystal structures represent only a small subset of the possible local minima.

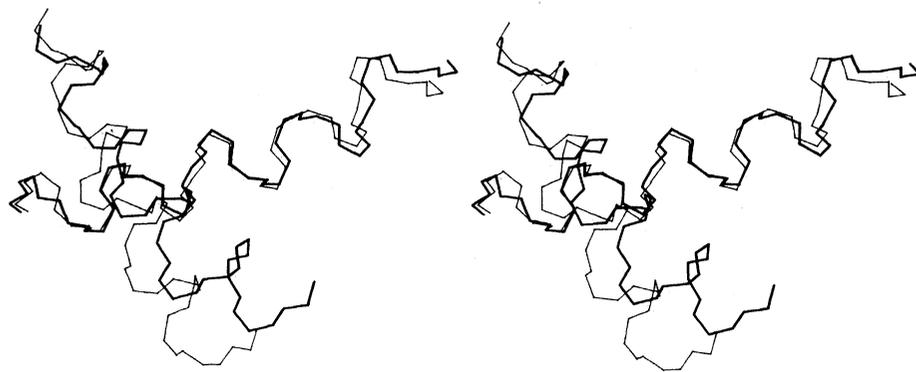
The complexity of the conformational space and the existence of multiple minima



**Fig. 4.** Minimized structures of myoglobin; stereo diagrams showing two superposed structures separated by a time interval of 100 psec along the trajectory.



**Fig. 5.** Averaged scaled difference matrices for  $C_\alpha$  atoms; the indices on the ordinate and abscissa designate the residue number. All values greater than 30 percent of the largest value are shown (most values are between 30 and 50 percent); positive differences are plotted above the diagonal, negative below it. (A) Structures sampled at 100-psec time interval; the secondary structural elements are indicated at the top of the figure; (B) structures sampled at 200-psec time interval.



**Fig. 6.** Helices A and H in two superposed minimized structures (25 and 300 psec); the main-chain atoms N,  $C_\alpha$ , C are shown.

with small energy differences suggest that a protein is a disordered system that may have glasslike properties at low temperatures (18, 19). We focus here on one striking property of spin glasses, namely, their ultrametricity (15, 29, 30). For a random spin system, it has been shown (30) that for a set of replica structures (each one is defined by a stable orientation of the spins), a similarity function can be defined such that the replicas fall into disjoint clusters, in which the overlap of all pairs is larger than a specified (but arbitrary) value; the size of the cluster depends on the choice of the overlap criterion. Such an ultrametric system is isomorphous with an evolutionary tree (15, 29). To apply this concept to the myoglobin simulation, we consider each of the minimized structures as a replica and use the rms difference between them as the overlap criterion. There is a rather sharp transition between the range ( $\text{rms} > 1.5 \text{ \AA}$ ) when all structures are disjoint, and the range ( $0 < \text{rms} < 1 \text{ \AA}$ ) when all the structures belong to the same cluster. In a very narrow region ( $1 < \text{rms} < 1.5 \text{ \AA}$ ), there are sets of disjoint clusters with more than one structure per cluster; for  $\text{rms} = 1.2 \text{ \AA}$ , there are clusters of size 1 (several) and one each of 2, 4, 7, and 9. Since the rms overlap criterion is arbitrary, we have tried others (for example, distance matrices) and have obtained corresponding results.

It appears that ultrametricity is not a useful concept when applied to the myoglobin simulation. What this means for actual myoglobin molecules or other proteins is not clear, since the simulation is short and the structures form a sequence deviating in time. Also, it is not known whether glasses [in contrast to the special model for a spin glass considered in (30)] are ultrametric.

Myoglobin at normal room temperatures samples a very large number of different minima that arise from the inhomogeneity of the system. This is expected to have important consequences for the interpretation of myoglobin function and, more generally, for the functions of other proteins, including enzymes. There are solidlike microdomains (the helices), whose main-chain structure is relatively rigid, and liquidlike regions (the loops and the side-chain clusters at interhelix contacts) that readjust as the helices move from one minimum to another. Since the minima have similar energies, myoglobin is expected to be glasslike at low temperatures. Freezing in of the liquidlike regions could result in a transition to the glassy state.

#### REFERENCES AND NOTES

1. M. Karplus and J. A. McCammon, *Annu. Rev. Biochem.* **53**, 263 (1983).
2. M. Karplus and J. N. Kushick, *Macromolecules* **14**, 325 (1981).
3. B. R. Brooks and M. Karplus, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 6571 (1983).

4. M. Levitt, C. Sanders, P. S. Stern, *J. Mol. Biol.* **181**, 423 (1985).
5. W. Bialek and R. F. Goldstein, *Biophys. J.* **48**, 1027 (1985).
6. R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder, I. C. Gunsalus, *Biochemistry* **14**, 5355 (1975).
7. H. Frauenfelder, G. A. Petsko, D. Ternoglou, *Nature (London)* **280**, 558 (1979).
8. R. M. Levy, D. Perahia, M. Karplus, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1349 (1982).
9. P. G. Debrunner and H. Frauenfelder, *Annu. Rev. Phys. Chem.* **33**, 283 (1983).
10. T. Ichiye, B. Olafson, S. Swaminathan, M. Karplus, *Biopolymers*, in press.
11. W. S. Bennett and R. Huber, *CRC Crit. Rev. Biochem.* **15**, 291 (1985).
12. C. B. Brooks, M. Karplus, B. M. Pettitt, *Adv. Chem. Phys.*, in press.
13. M. Agmon and J. J. Hopfield, *J. Chem. Phys.* **79**, 2042 (1983).
14. A. Ansari *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 5000 (1985).
15. G. Toulouse, *Helv. Phys. Acta* **57**, 459 (1984).
16. F. H. Stillinger and T. A. Weber, *Phys. Rev.* **A25**, 978 (1982).
17. ———, *Science* **225**, 983 (1984).
18. D. L. Stein, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3670 (1985).
19. J. M. Ziman, *Models of Disorder* (Cambridge Univ. Press, London, 1979), section 7.2.
20. A. M. Lesk and C. Chothia, *J. Mol. Biol.* **136**, 225 (1980).
21. R. M. Levy *et al.*, *Biophys. J.* **48**, 509 (1985).
22. B. R. Brooks *et al.*, *J. Comp. Chem.* **4**, 187 (1983).
23. S. Swaminathan, T. Ichiye, W. van Gunsteren, M. Karplus, *Biochemistry* **21**, 5230 (1982).
24. G. P. Singh *et al.*, *Z. Phys.* **B55**, 23 (1984).
25. F. Parak, E. W. Knapp, D. Kucheida, *J. Mol. Biol.* **161**, 177 (1982).
26. T. F. Havel, I. D. Kuntz, G. M. Crippen, *Bull. Math. Biol.* **45**, 665 (1983).
27. J. Kuriyan, M. Karplus, G. A. Petsko, *J. Mol. Biol.*, in press.
28. J. L. Smith *et al.*, *Biochemistry* **25**, 5018 (1986).
29. M. Mezard, G. Parisi, N. Sourlas, G. Toulouse, M. Virasoro, *Phys. Rev. Lett.* **52**, 1156 (1984).
30. A. T. Ogielski and D. L. Stein, *ibid.* **55**, 1634 (1985).
31. We thank G. S. Dubey, J. W. Keepers, R. M. Levy, R. P. Sheridan, and S. Swaminathan for their essential contribution to the myoglobin simulation and A. Beyer, J. Kuriyan, and N. Summer for very helpful discussions and comments. This work was supported in part by a grant from the National Institutes of Health.

7 July 1986; accepted 12 November 1986

## Construction of a Novel Oncogene Based on Synthetic Sequences Encoding Epidermal Growth Factor

D. F. STERN, D. L. HARE, M. A. CECCHINI, R. A. WEINBERG

**The autocrine model postulates that constitutive release of a mitogenic growth factor can lead to uncontrolled proliferation and cell transformation. A synthetic polynucleotide encoding epidermal growth factor conferred a tumorigenic phenotype on cells. These cells were transformed through the action of an autocrine circuit having an extracellular component.**

**N**EOPLASTIC TRANSFORMATION OF cells by oncogenes and carcinogens results in the elaboration of growth factors (GF's). These include transforming growth factors (TGF's), which are able to induce some aspects of the transformed phenotype in nontransformed cells (1). Since such factors may affect the behavior of the cell that has released them, these findings suggested an autocrine mechanism of transformation whereby production of TGF's by a cell leads to uncontrolled stimulation of its own growth (2).

The role of TGF's in transformation has been difficult to assess because transformed cells produce a number of factors that interact in complex ways [for example, see (3)] and because the properties of TGF's have been studied primarily in culture. Furthermore, it remained unclear whether induction of transformation is a special property of TGF's, or whether constitutive production of any mitogen by cells that can respond to this mitogen is sufficient to induce transformation. To address these questions, we have used an expression vector to induce constitutive production and secretion of human epidermal growth factor (EGF). EGF

was chosen for these studies because the factor and its receptor are well characterized (4) and because transformed cells frequently produce TGF- $\alpha$ , a factor that binds to the EGF receptor and has nearly identical physiological properties to those of EGF (5).

The EGF expression vector, designated pUCDS3 (6), consists of a murine leukemia virus long terminal repeat (LTR) that contains promoter/enhancer sequences, an SV40 polyadenylation site, sequences encoding a mouse immunoglobulin heavy-chain signal peptide (7) (to cause secretion of the translation product), and chemically synthesized human EGF-encoding sequences (originally designed for translation in *Escherichia coli*) (8) (Fig. 1). The immunoglobulin (Ig) signal peptide-encoding sequence was mutagenized to create an Eco RI cleavage site (Fig. 1C). To reconstruct the signal peptidase cleavage site, the 5' end of a 175-bp synthetic gene encoding EGF

D. F. Stern and R. A. Weinberg, Whitehead Institute for Biomedical Research, Cambridge, MA 02142, and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.  
D. L. Hare and M. A. Cecchini, Amgen Development, Inc., Boulder, CO 80301.