The Elongation-Termination Decision in Transcription

PETER H. VON HIPPEL AND THOMAS D. YAGER

At any template position, the decision to extend the transcript by one residue or to release the nascent RNA represents a kinetic competition between elongation and termination pathways. This competition is discussed in terms of alternative Eyring transition state barriers; changes in termination efficiency correspond to small changes in the relative heights of these barriers. Elongation complexes are stable at nonterminator positions; a model is presented to explain the destabilization of these complexes at intrinsic termination sites. Functionally analogous effects can operate at rho-dependent terminators. Mechanisms for modulation of termination efficiency by regulatory proteins are described.

TRANSCRIPTION IN *ESCHERICHIA COLI* PROCEEDS IN THREE phases. First, an initiation complex containing RNA polymerase forms at the promoter of the gene to be transcribed, and the first few residues of RNA are synthesized. This initiation complex is relatively unstable and prone to dissociate with the release of short (abortive) RNA products. At the end of the initiation phase, RNA polymerase loses contact with the promoter, the sigma specificity subunit is released, core polymerase (which lacks sigma factor) undergoes a major conformational rearrangement, and the stable elongation complex is formed [reviewed in (*1*)].

The elongation complex then proceeds along the DNA template, extending the nascent RNA in a totally processive manner. RNA synthesis is highly regulated in this phase, both kinetically and thermodynamically, by sequence-dependent interactions of the core polymerase with the DNA template and the RNA transcript and by interaction of the elongation complex with protein factors. Consequently, progression of the elongation complex along the DNA is irregular, with both the rate of RNA synthesis and the stability of the complex varying significantly with template position.

The elongation complex has two choices at a given template position I: it can extend the nascent RNA chain by one nucleotide to form a transcript that is $I + 1$ residues in length, or it can be induced by features of the DNA or RNA sequence (perhaps assisted by specific protein termination factors) to form an unstable termination complex. This termination complex has a significant probability of dissociating from the DNA template, resulting in the release of a terminated RNA transcript of length I.

Termination is observed only at two specific types of DNA sites. Intrinsic terminators induce the complex to switch from the elongation mode to the termination mode by a mechanism that depends only on nucleic acid sequence (although protein factors may modulate the effect). Rho-dependent terminators carry signals in the DNA (and perhaps in the RNA) that cause the elongation complex to pause at these sites. However, actual RNA release depends on interaction of the paused complex with transcription termination factor rho.

This work was started while P. H. von Hippel and T. D. Yager were at the Institute of Molecular Biology and Department of Chemistry, University of Oregon, Eugene, OR 97403. T. D. Yager has since moved to the National Science Foundation Center for Molecular Biotechnology, Division of Biology, California Institute of Technology, Pasadena, CA 91125.

Alternative Transition State Barriers

A competitive kinetic model of the elongation-termination decision (Fig. 1) can be proposed independently of any particular structural representation of the transcription complex. In such a model, the elongation and termination processes are characterized by specific overall rate constants at each template position (*2*). The magnitudes of these rate constants are expected to depend on DNA and RNA sequences and on interactions of the transcription complex with extrinsic protein factors.

The average step time for the addition of the next nucleotide residue is ~30 ms at a typical nonterminator site in vitro (*3*). Using the Eyring transition state theory, one can calculate that this step time corresponds to an activation barrier height to elongation of ~+16 kcal/mol (Fig. 1A) (*2*). In contrast, at such a position the barrier height to termination (dissociation and RNA release) is probably >+30 kcal/mol, as suggested by the pronounced stability of elongation complexes that have been artificially stalled by nucleotide triphosphate depletion (*4*). Clearly at nonterminator sites the barrier to elongation is much lower than the barrier to termination, and thus transcription should proceed with an infinitesimal probability of spontaneous RNA release.

Regulation of Termination Efficiency

In contrast, the elongation complex is destabilized at terminator positions (Fig. 1B). At these loci, RNA release occurs in 1 to 10 s or less (*4, 5*). If the average dwell time (the length of time the complex pauses at I) for the elongation complex at these positions is also ~1 to 10 s, the heights of the activation barriers to elongation and to termination will be approximately equal at ~+18 kcal/mol each. Consequently, minor changes in the relative barrier heights (and thus in the relative rates of the two competing processes) should cause major changes in termination efficiency.

In vitro transcription reactions are typically conducted with a DNA template that contains a single promoter and a single terminator. Under these conditions, a fraction of the population of elongating RNA transcripts is released at the terminator at position I; this fraction of the transcripts (RNA_T) is detected as a specific band on a polyacrylamide gel. The remaining fraction of the transcripts (RNA_{runoff}) is extended to the end of the DNA and is detected as a second specific band. The efficiency of termination (TE) at I is defined as:

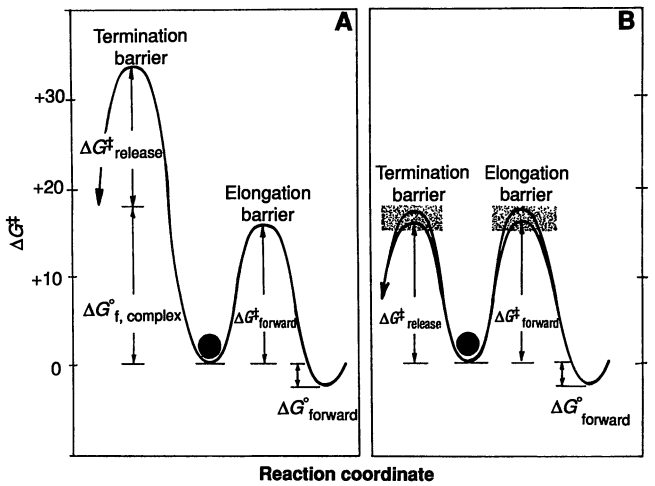

Fig. 1. Schematic diagram of the relative activation barrier heights for elongation and termination at a typical elongation site (**A**) and a typical termination site (**B**). Barrier heights corresponding to a tenfold change in pause time (or a 1.4 kcal/mol change in barrier height) at the terminator are shown as alternative peaks in (B). Stippled areas indicate changes in barrier heights that correspond to tenfold changes in *TE*. [Reproduced from (*2*)] $\Delta G^{\ddagger}_{forward}$ and $\Delta G^{\ddagger}_{release}$ represent the heights of the activation barriers to transcript elongation and release at template position *I*, respectively. $\Delta G^{\circ}_{forward}$ represents the thermodynamic free energy of extending the transcript from *I* to *I* + 1 (*15*).

$$TE_I = RNA_I \div (RNA_I + RNA_{runoff}) \quad (1)$$

In a competitive kinetic (Eyring) model as presented in Fig. 1, the termination efficiency of a transcription complex located at *I* will be

$$TE_I = [1 + e^{-\Delta(\Delta G^{\ddagger})/RT}]^{-1} \quad (2)$$

where $\Delta G^{\ddagger}$ is the activation free energy, $\Delta(\Delta G^{\ddagger})$ is the difference in height between the two activation barriers of Fig. 1 (*2*), *R* is the gas constant, and *T* is the temperature. Thus, TE_I ranges from 0 to 1 at each template position as a function of the difference in the (position-dependent) heights of the activation barriers to elongation and termination. We emphasize that Eq. 2 is based on general kinetic principles and is independent of any particular structural model of the elongation or termination complex.

The theoretical curves of Fig. 2 represent TE_I as a function of $\Delta(\Delta G^{\ddagger})$ according to Eq. 2. These curves predict that termination efficiency should change from >99% to <1% in response to a change in $\Delta(\Delta G^{\ddagger})$ of only 3 to 4 kcal/mol. This magnitude of change in relative barrier heights is indicated by the stippled areas of Fig. 1B.

The competitive kinetic model makes two predictions. First, RNA release should be possible only at sites where the elongation complex is destabilized to such an extent that the heights of the elongation and termination barriers are approximately equal. Because of the great discrepancy in barrier heights at nonterminator template positions (Fig. 1A), the position of terminators along the template should be strongly determined, and the elongation-termination decision should have the character of a binary switch. Second, termination efficiencies should be easily regulated at termination sites, because relatively small changes in either kinetic or thermodynamic components of the activation free energy barriers will produce large changes in TE_I as a consequence of the exponential form of Eq. 2.

This model suggests two questions. First, how do certain DNA sequences destabilize the elongation complex, resulting in the approximate equalization of the elongation and termination barrier heights at terminators? Second, what are the intrinsic (sequence-dependent) effects or the extrinsic (regulatory protein–dependent) effects that determine the efficiency of termination at such termination sites?

Structural Models and Thermodynamic Consequences

These questions can be addressed in terms of the specific structural model of the elongation complex that is presented in Fig. 3. The evidence that supports this consensus model has been summarized (*6, 7*). We have defined a thermodynamic stability function (*7*) that describes the standard free energy of formation of the elongation complex ($\Delta G^{\circ}_{f,\ complex}$) at any template position in terms of three components:

$$\Delta G^{\circ}_{f,\ complex} = \Delta G^{\circ}_{f,\ DNA\text{-}bubble} + \Delta G^{\circ}_{f,\ RNA\text{-}DNA\ hybrid} + \Delta G^{\circ}_{f,\ pol\text{-}binding} \quad (3)$$

Here $\Delta G^{\circ}_{f,\ DNA\text{-}bubble}$ is the (unfavorable) standard free energy of formation of the open DNA bubble from intact duplex DNA, $\Delta G^{\circ}_{f,\ RNA\text{-}DNA\ hybrid}$ is the (favorable) standard free energy of formation of the RNA-DNA hybrid (within the open DNA bubble) from single-stranded DNA and RNA chains, and $\Delta G^{\circ}_{f,\ pol\text{-}binding}$ is the (favorable) standard free energy of binding of the polymerase to the nucleic acid components of the ternary complex. This thermodynamic formulation is general although the magnitudes of the free energy terms in Eq. 3 will depend on the details (such as the size of the DNA bubble and the length and position of the DNA-RNA hybrid) of the structural model that is assumed. We note that a model differing somewhat from that of Fig. 3 (primarily in the length of the DNA-RNA hybrid) has also been proposed (*8*).

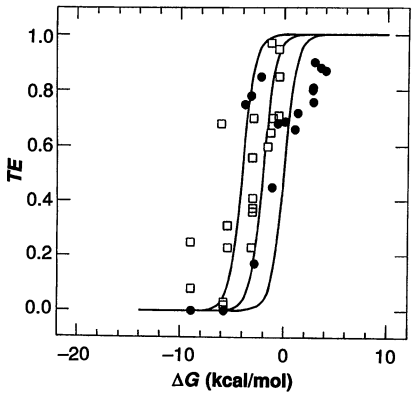

Fig. 2. Measured (in vitro) efficiencies of intrinsic terminators (*TE*) plotted as a function of the calculated value of $\Delta G^{\circ}_{f,\ complex,\ max}$ at the termination positions; (●), *thr* attenuator variants; (□), *trp* attenuator variants. Theoretical curves of *TE* versus $\Delta(\Delta G^{\ddagger})$ are superimposed. [Reproduced from (*2*)]

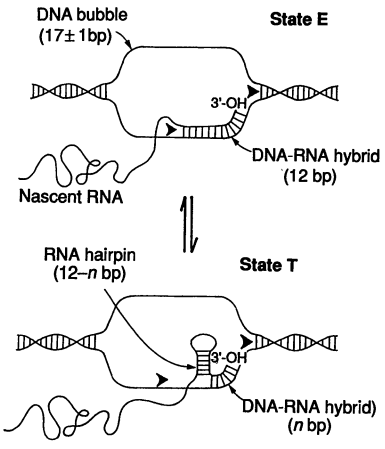

Fig. 3. Structural models of the consensus elongation complex (state E) and of the termination complex at an intrinsic termination site (state T); *n*, the number of base pairs remaining in the DNA-RNA hybrid. [Modified with permission from (*7*), American Chemical Society]

We have estimated $\Delta G°_{f,\ complex}$ for a number of *E. coli* operons (*6, 7*) using the specific model of the elongation complex diagramed in Fig. 3, thermodynamic data that establish the stability of DNA-DNA, RNA-DNA, and RNA-RNA sequences, and an argument that sets $\Delta G°_{f,\ pol\text{-}binding}$ at -30 ± 3 kcal/mol and suggests that this term is approximately independent of template position. Our calculations suggest that the net thermodynamic stability of the elongation complex averages ~ -18 kcal/mol at nonterminator sites. This thermodynamic stability factor adds to the height of the activation barrier that must be overcome to permit RNA release at nonterminator sites and thus may account for the great disparity in barrier heights to elongation and termination at these sites (Fig. 1A).

The consensus structural model of the elongation complex (Fig. 3) can also explain the abrupt loss of transcription complex stability that accompanies formation of the termination complex at intrinsic terminators (7). We and others have proposed that, as the elongation complex approaches an intrinsic termination site, RNA residues that can form a GC-rich RNA hairpin in the nascent RNA transcript are synthesized. At a critical template position, the stability of the potential RNA hairpin within the growing transcript exceeds that of the corresponding portion of the RNA-DNA hybrid. The hairpin then forms cooperatively and a portion of the (~12 bp) hybrid is displaced. The shortened RNA-DNA hybrid that remains consists primarily of uridine-adenine base pairs, which are particularly unstable (*9*). Thus, the stabilizing effect of the $\Delta G°_{f,\ RNA\text{-}DNA\ hybrid}$ term of Eq. 3 is abruptly reduced (over one to two nucleotides along the DNA template) as an intrinsic terminator is encountered. As a consequence, $\Delta G°_{f,\ complex}$ is decreased to ~0 kcal/mol at such a terminator (7). This destabilization of the RNA-DNA hybrid over one or two residues along the DNA template can explain the exactness with which the positions of intrinsic termination sites are specified (*6, 7*). We emphasize that any alternative model of the transcription complex and of the structural changes that accompany the elongation-termination transition must explain this discontinuous (binary switch) stability behavior.

In Fig. 2, we plot in vitro *TE* data as a function of the maximum value of the free energy of formation of the transcription complex, $\Delta G°_{f,\ complex,\ max}$, calculated at known termination sites with Eq. 3. The *TE* values shown were obtained from the literature (*2*) and were measured with DNA fragments that contain nucleotide sequence variants of the *Escherichia coli trp* and *thr* attenuators; these variants differ from the wild type at only one or a few nucleotide positions. In Fig. 2, these data have been superimposed on theoretical curves of *TE* versus $\Delta(\Delta G^{\ddagger})$ calculated with Eq. 2. Three theoretical curves, offset from one another by 2 kcal/mol along the abscissa, indicate the magnitude of the scatter of the experimental data. Within a range of 2 to 3 kcal/mol, the data points for a given terminator fall within the distribution predicted by the competitive kinetic scheme.

It is clear, however, that in Fig. 2 many of the data points within a given set of terminator variants deviate significantly from any one theoretical curve. These deviations could in principle be attributed to random error in the experimental measurements of *TE*. It is more likely, however, that they reflect alterations in specific interactions within the transcription complex. To show how the deviation of an experimental point from a given theoretical curve might arise, we consider the following example.

Suppose that a particular wild-type terminator displays a *TE* of 0.5, and a specific nucleotide substitution in this terminator alters the observed *TE* to 0.16 (an antitermination phenotype). Suppose also that this difference is reproducible and significantly larger than the random error associated with the measurement of *TE*. We calculate from Eq. 2 that this difference in *TE* corresponds to a change in $\Delta(\Delta G^{\ddagger})$ of -1 kcal/mol. At least three experimentally testable interpretations of this change in *TE* can be offered.

1) The observed change might reflect a -1 kcal/mol increase in the stability of the nucleic acid interactions within the transcription complex, that is, a change of -1 kcal/mol in the sum of the $\Delta G°_{f,\ DNA\text{-}bubble}$ and $\Delta G°_{f,\ RNA\text{-}DNA\ hybrid}$ terms of Eq. 3. This would increase the height of the termination barrier by 1 kcal/mol relative to the height of the elongation barrier. Within the limits of accuracy of the DNA-DNA, DNA-RNA, and RNA-RNA thermodynamic data, both *TE* versus $\Delta G°_{f,\ complex,\ max}$ points would be expected to fall on the same theoretical *TE* versus $\Delta(\Delta G^{\ddagger})$ curve in this case (Fig. 2), confirming that the effect of the nucleotide change on *TE* is due to a change in the $\Delta G°_{f,\ DNA\text{-}bubble}$ or the $\Delta G°_{f,}$ RNA-DNA hybrid term of Eq. 1. A stability change of this sort should be directly measurable.

2) The *TE* change could also correspond to a -1 kcal/mol change in the binding affinity of the polymerase for the nucleic acid components of the transcription complex (that is, to a change in $\Delta G°_{f,\ pol\text{-}binding}$). In this case, the two experimental *TE* versus $\Delta G°_{f,\ complex,\ max}$ points would not fall on the same theoretical *TE* versus $\Delta(\Delta G^{\ddagger})$ curve. This stability change should also be directly measurable.

3) The observed *TE* change could reflect a roughly fivefold decrease in pause time at the terminator, corresponding to a 1 kcal/mol decrease in height of the elongation barrier. In this case, the two experimental *TE* versus $\Delta G°_{f,\ complex,\ max}$ points would also not fall on the same theoretical *TE* versus $\Delta(\Delta G^{\ddagger})$ curve. Such a change in dwell time should also be directly measurable.

In connection with the third interpretation, it has been shown that binding of antitermination proteins Q (*10*) and N (*11*) of phage λ to certain elongation complexes does decrease pause times along the template. This observation suggests that the competitive kinetic model can be used to interpret the modulation of the efficiencies (but not of the positions) of intrinsic terminators by at least some regulatory proteins. Modulation can be achieved either kinetically or thermodynamically, as indicated in the described example.

Finally, we note that the model can also be used to interpret some aspects of RNA release at rho-dependent termination sites. Sequence analysis of the 3′-termini of rho-terminated transcripts shows that (unlike at intrinsic terminators) destabilization of the elongation complex by RNA hairpin formation is not likely. However, a competitive kinetic approach may again apply, because the transcription complex has been shown to pause at rho-dependent terminators for times up to several 100-fold longer than typical elongation step times (*12*). Such a pause raises the height of the barrier to elongation at these sites by a known amount and thus favors termination to a predictable extent (Eq. 2). Measurement of rho-dependent termination efficiencies with slow polymerase mutants are consistent with this interpretation (*13*). The actual release by rho of the nascent RNA chain from the transcription complex may occur by means of a thermodynamic mechanism, because rho is an adenosine triphosphate–dependent RNA-DNA helicase (*14*) and thus should destabilize the $\Delta G°_{f,\ RNA\text{-}DNA\ hybrid}$ component of the termination complex.

Conclusions and Perspectives

These findings suggest that the competitive kinetic model we have summarized can provide a general framework for the analysis of transcript termination induced by nucleotide sequence and by interaction with regulatory proteins. The termination process may reflect kinetic effects, thermodynamic effects, or both. The magnitudes of these effects can be predicted and subjected to experimental test. Finally, we note that these ideas should also be applicable to the

analysis of the elongation-termination decision in other prokaryotic and eukaryotic organisms, especially as details of structure and interactions in the elongation complexes of these systems become available.

REFERENCES AND NOTES

1. S. C. Gill, T. D. Yager, P. H. von Hippel, *Biophys. Chem.* **37**, 239 (1990).
2. P. H. von Hippel and T. D. Yager, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 2307 (1991).
3. G. Rhodes and M. J. Chamberlin, *J. Biol. Chem.* **249**, 6675 (1974).
4. K. M. Arndt and M. J. Chamberlin, *J. Mol. Biol.* **213**, 79 (1990).
5. C. Andrews and J. P. Richardson, *J. Biol. Chem.* **260**, 5826 (1985).
6. T. D. Yager and P. H. von Hippel, in Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology*, F. C. Neidhardt, Ed. (American Society of Microbiology, Washington, DC, 1987), pp. 1241–1275.
7. ———, *Biochemistry* **30**, 1097 (1991).
8. G. A. Rice, C. M. Kane, M. J. Chamberlin, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4245 (1991).
9. F. H. Martin and I. Tinoco, Jr., *Nucleic Acids Res.* **8**, 2295 (1980).
10. X. Yang and J. W. Roberts, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5301 (1989).
11. W. A. Whelan and A. Das, unpublished results.
12. W. D. Morgan, D. G. Bear, P. H. von Hippel, *J. Biol. Chem.* **258**, 9565 (1983).
13. D. Jin, R. R. Burgess, J. P. Richardson, C. Gross, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
14. C. A. Brennan, A. J. Dombroski, T. Platt, *Cell* **48**, 945 (1987).
15. T. D. Yager, J. Korte, P. H. von Hippel, unpublished data.

16. We thank many colleagues for discussions and A. Das, C. Gross, and D. Jin for providing unpublished information. Supported in part by USPHS research grants GM-15792 and GM-29158 (to P.H.v.H.) and by a grant to the University of Oregon from the Lucille P. Markey Charitable Trust. P.H.v.H. is an American Cancer Society Research Professor of Chemistry.

Valuing the Health Benefits of Clean Air

JANE V. HALL, ARTHUR M. WINER, MICHAEL T. KLEINMAN, FREDERICK W. LURMANN, VICTOR BRAJER, STEVEN D. COLOME

An assessment of health effects due to ozone and particulate matter (PM_{10}) suggests that each of the 12 million residents of the South Coast Air Basin of California experiences ozone-related symptoms on an average of up to 17 days each year and faces an increased risk of death in any year of 1/10,000 as a result of elevated PM_{10} exposure. The estimated annual economic value of avoiding these effects is nearly $10 billion. Attaining air pollution standards may save 1600 lives a year in the region.

IN THE SOUTH COAST AIR BASIN OF CALIFORNIA (THE BASIN), air quality remains the worst in the nation with respect to particulate matter (PM_{10}), ozone (O_3), and nitrogen dioxide (NO_2). Attaining national ambient air quality standards (NAAQS) (*1*) in the basin will require pervasive, technology-forcing emission controls in addition to changes in life-style. Efforts to determine whether these controls are worthwhile depend on studies that estimate how much health is affected and the value of avoiding these effects.

Executive Order 12291 signed by President Reagan in 1981 requires benefit-cost comparisons for new federal rules, including air quality regulations. The impetus behind this approach—determining how to use limited dollars to provide the greatest good—has led regional governments to pose questions about the worth of pollution-control programs. Estimating the level of benefits resulting from NAAQS attainment is a crucial part of this equation, but remains the more challenging task.

Our approach and resulting estimates of health benefits are somewhat different from those in earlier studies (*2–4*). For O_3-related effects, we included different symptoms from those of earlier workers, assumed that 8 parts per hundred million (pphm) over a 1-hour period was the threshold for O_3-related effects, used more clinical data to derive and validate dose-response functions, and used a fuller characterization of population exposure and dose. For PM_{10}, the major difference is the value of life used to estimate dollar benefits. In the absence of any clear evidence on the age distribution of premature deaths, we used a value from the middle of the generally accepted range rather than the low end.

J. V. Hall and V. Brajer are in the Department of Economics, California State University, Fullerton, CA 92634. A. M. Winer is in the Environmental Science and Engineering Program, Department of Environmental Health Sciences, School of Public Health, University of California, Los Angeles, CA 90024. M. T. Kleinman is in the Department of Community and Environmental Medicine, Air Pollution Health Effects Laboratory, University of California, Irvine, CA 92715. F. W. Lurmann is with Sonoma Technology, Incorporated, Santa Rosa, CA 95403. S. D. Colome is with Integrated Environmental Services, Irvine, CA 92715.

Exposure and Dose Estimation

To generate useful estimates, human exposure modeling must effectively capture exposure and dose received by the entire population (*5*). This process can be difficult but is important because people's susceptibility to health effects varies according to their age and the kinds and times of their activities.

We characterized exposure and dose as fully as possible with the Regional Human Exposure (REHEX) model (*6*), which estimates a population's typical outdoor, indoor, and in-vehicle exposures (*7*) by time of day. Critical to this model is a statistical estimation of the basin population's spectrum of time-activity patterns (*8*).

For each of nine demographic groups, we generated approximately 1000 time-activity patterns (*9*) consisting of typical 15-min intervals spent in a series of microenvironments where pollution levels remained spatially uniform (*5*). We then subdivided each demographic group into time-activity cohorts that followed specific patterns and, using distributions from epidemiological and transportation time-use studies (*10*), estimated an expected frequency of occurrence for each of these patterns. We estimated the concentrations of pollutants to which each group was exposed by dividing the basin into 32 districts corresponding to the ambient air monitoring network of the South Coast Air Quality Management District (SCAQMD). Because of the severity and pervasiveness of NAAQS violations for O_3 and PM_{10} in the basin, we concentrated on these compounds. Hourly O_3 data and 24-hour-average PM_{10} data were used. Limited simultaneous indoor and outdoor measurements were used to determine levels of indoor O_3 resulting from penetration of