saturation at which nucleation occurs is determined largely by the size of the nanodroplets present (10).

Could the precritical clusters observed by Gebauer et al. also be the result of stabilization of calcium carbonate clusters by another species present as an impurity? This mechanism would provide a basis for stabilizing precritical clusters in a free-energy minimum and does not contradict classical nucleation theory. Such impurities are ubiquitous and virtually impossible to eliminate from any solution. The results of Gebauer et al. may thus reflect the mechanism of nucleation of calcium carbonate in “real” systems. Nucleation could then occur by coalescence of the precritical clusters to give ACC, which will subsequently crystallize to a more stable crystalline polymorph. The latter mechanism is consistent with the observations of Gebauer et al., who show that ACC is the first phase precipitated after nucleation.

The idea that nucleation of calcium carbonate may proceed via an aggregation mechanism is highly topical. The past decade has seen great progress in understanding crystallization processes, and it is now well recognized that single-crystal growth (as distinct from nucleation) often occurs via the aggregation of small precursor units rather than by addition of ions or molecules to a nucleus (11). Cluster species have also been observed before nucleation in saturated solutions of compounds such as sodium chloride (2), urea (12), and glycine (3), and there have been suggestions that clustering can determine which polymorph is formed (13). However, none of these even remotely approach the size or stability of the clusters observed by Gebauer et al. Further investigation of precritical clusters and their role in the crystallization of calcium carbonate, and indeed other compounds, is eagerly anticipated.

References

PERSPECTIVES

DEVELOPMENTAL BIOLOGY

From Genetic Association to Genetic Switch

Alan M. Michelson

A

Deciphering the sequence of the human genome and the subsequent cataloging of common human DNA sequence variation marked a paradigm shift in human genetics. These resources, together with advances in cost-effective genotyping technologies, enabled the design of genome-wide association studies for the unbiased discovery of commonly occurring DNA sequence variations called single-nucleotide polymorphisms (SNPs) that are preferentially associated with a disease or other clinical trait (1). Although genome-wide association studies have uncovered disease-associated SNPs, identifying actual disease-causing variants—and gaining deep insights into how those variants generate the underlying molecular pathophysiology—have so far yielded only modest results. This has led to criticisms of the genome-wide association approach for investigating the etiologies of common diseases (2). However, this assessment may be premature. On page 1839 of this issue, Sankaran et al. (3) show how genome-wide association findings can lead to a detailed understanding of disease mechanisms and be used to ascertain novel therapeutic targets.

Previous genome-wide association studies conducted in independent populations have identified SNPs in three chromosomal loci that are associated with varying expression levels of human fetal hemoglobin (HbF) (4–6). HbF is a clinically important quantitative trait because elevated concentrations correlate inversely with HbF production; low mRNA expression corresponds to homozygosity for the allele associated with low HbF production; low mRNA expression of BCL11A encodes a protein that represses transcription in the B lymphoid lineage (8).

Sankaran et al. hypothesized that BCL11A might repress expression of the γ-globin gene, with expression or activity of this repressor correlating inversely with HbF production both during normal development and in individuals of different genotypes at the BCL11A locus. They first determined that BCL11A is expressed as two long isoforms, encoded by alternatively spliced messenger RNAs (mRNAs), in primary adult erythroblasts. By contrast, only shorter variants of BCL11A are found in human embryonic erythroleukemia cells and in primary human fetal liver cells, both of which express high amounts of HbF. Moreover, the genotype at the BCL11A SNP that affects HbF production influences expression of mRNAs encoding the long isoforms in lymphoblastoid cell lines: High expression of BCL11A mRNA corresponds to homozygosity for the allele associated with low HbF production; low mRNA expression corresponds to homozygosity for the allele associated with high HbF production; and SNP heterozygotes express intermediate amounts of mRNA (see the figure). If the association between the BCL11A SNP and the expression level of this gene in lymphoblas-
it functions in this capacity. Higher-resolution studies of BCL11A chromatin occupancy, functional characterization of the putative cis-regulatory elements containing these binding sites, and sequencing of the entire region of BCL11A that is in linkage disequilibrium with the SNP used to discover its relevance to HbF expression are needed to address these questions.

The findings of Sankaran et al. also have potential consequences for developing new therapies to treat sickle cell disease and other hemoglobinopathies. The inverse correlation between BCL11A and HbF expression, combined with the known ameliorative effect of HbF on the pathophysiology of sickle cell disease and β-thalassemia, suggests that inhibition of BCL11A expression or function could be an effective treatment for these disorders. The study also illustrates that, when experiments are appropriately designed, the initial findings of genome-wide association studies can be successfully followed up at a functional level.

References
3. V. G. Sankaran et al., Science 322, 1839 (2008); published online 4 December 2008 (10.1126/science.1165409).

PERSPECTIVES

TRANSCRIPTION

Gene Expression—Where to Start?

Stephen Buratowski

Transcription just got noisier with the discovery of short RNAs that are synthesized at or near DNA regions that also initiate full-length RNAs.

To convert the encoded genetic information from eukaryotic DNA into proteins, base sequences of genes are first transcribed into RNA by RNA polymerase II. To produce functional RNA molecules, dozens of accessory factors are needed to define the proper locations for RNA polymerase II to begin and end transcription. Although we have some basic knowledge about how these factors work, it is still not possible to take a eukaryotic genome sequence and accurately predict what RNA species it will produce. Recent efforts to map and sequence “transcriptomes” have only increased the challenge by revealing a much more complex set of RNAs than expected, including many that do not produce proteins. The latest surprise, described in four papers in this issue (1–4), is a new class of transcripts that initiate near the expected transcription start sites upstream of protein-encoding sequences. However, these RNAs are short, present at low abundance, and often occur in the direction opposite to that of the protein-coding region (see the figure). It remains to be seen whether these RNAs have a function, but their existence challenges our simplistic models about how the DNA sequences known as “promoters” define transcription start sites.

In current textbook models, promoters comprise two interacting parts. Basal promoter elements bind accessory transcription initiation factors that position RNA polymerase II in the right place and direction. Enhancer elements bind regulatory factors...
From Genetic Association to Genetic Switch
Alan M. Michelson

Science 322 (5909), 1803-1804.
DOI: 10.1126/science.1169216

ARTICLE TOOLS  http://science.sciencemag.org/content/322/5909/1803

RELATED CONTENT  http://science.sciencemag.org/content/sci/322/5909/1839.full

REFERENCES  This article cites 11 articles, 5 of which you can access for free
  http://science.sciencemag.org/content/322/5909/1803#BIBL

PERMISSIONS  http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service