Colloid Transport in the Subsurface: Past, Present, and Future Challenges

John F. McCarthy* and Larry D. McKay

ABSTRACT

This paper attempts to introduce the work described in this special section on colloid transport within a more general perspective of the evolution of our understanding of the importance of colloids in subsurface systems. The focus will be on the transport of colloidal particles in natural (i.e., chemically and physically heterogeneous) geological settings because the complexity imposed by these situations represents the greatest challenge to current and future understanding. Great progress has been made in addressing many of the key questions related to colloid transport. However, as in most areas of science, increased knowledge also serves to reveal new and more complex challenges that must be addressed.

In evaluating progress in this area, it is worth recalling that the nature of our interests in particle transport in the subsurface has changed dramatically in the past few decades. Initial interest focused on microbial contaminants in aquifers, then expanded to include the influence of colloids on soil development, permeability of oil and gas reservoirs, transport of contaminants attached to colloids, and introduction of microorganisms for remediation of contaminated subsurface materials. Our understanding of even the most basic issues, such as the appropriate methods for obtaining a representative sample of groundwater that includes mobile colloids while avoiding colloidal artifacts, has changed substantially. Progress was quickly made in some areas, such as the development of a sense of the distribution of subsurface colloids and the processes that controlled their nature and abundance. However, even as understanding of groundwater colloids progressed, it became clear that theoretical understanding was not always successful in describing behavior in natural systems. While most work to date has focused on colloid transport in saturated systems, increased attention is being directed at problems in vadose transport. The vadose zone represents the critical connection between shallow contaminant sources and migrating groundwater. The complexities related to dealing with water, solute, and colloid transport in partially saturated systems is a challenge that will lead us well into the future.

RECOGNIZING THE PROBLEMS

Interest in colloid behavior in porous media is not new. Early studies on colloid mobilization and transport focused on microbial contamination of aquifers and water supply wells. Many of the approaches for dealing with microbial contamination, such as establishing minimum setback distances between wells and septic fields (onsite wastewater disposal systems) for individual homes, were based on empirical evidence of contaminant filtration and are still in use today (TDEC, 2002). Other studies focused on the role of clay migration from surface to deeper horizons in soil diagenesis (Jenny and Smith, 1935; Thorp et al., 1957). Mechanistic studies of colloid transport originally focused on issues of wastewater filtration and controlling aquifer permeability. Yao et al. (1971) described colloid removal for water filtration in terms of two rate-limiting steps: (i) the physical processes of diffusion, interception, and gravitation settling that result in collisions between colloids and grains and (ii) the chemical factors controlling the interaction forces that result in attachment of the colloid to the grain surface. Many early studies on colloid behavior in geological media focused on the effects of aqueous chemical conditions on colloid dispersion and the permeability of geological formations in an effort to improve oil and gas recovery (Khilar and Fogler, 1984; Cerda 1987; Kia et al., 1987). An early field-scale manipulation of colloid dispersion resulted from concern about drinking water supply at an artificial recharge site in California. High turbidity in the groundwater was attributed to the low ionic strength recharge water, and application of gypsum (calcium sulfate) at the recharge area prevented release of clays and alleviated problems with groundwater turbidity (Nightingale and Bianchi, 1977).

Recognition that colloids might be important in facilitating transport of other contaminants arose from studies demonstrating that strongly sorbing contaminants could travel much further and faster than would be anticipated from traditional solute transport models based on batch sorption data. For example, at the Nevada Test Site, surface-reactive radionuclides were observed in groundwater outside nuclear detonation cavities (Coles and Ramsbottom, 1982; Buddemeier and Hunt, 1988; Kersting et al., 1999). Likewise, enhanced transport of radionuclides at Chalk River Nuclear Laboratories in Canada (Walton and Merritt, 1980; Champ et al., 1984), Oak Ridge National Laboratory in Tennessee (McCarthy et al., 1998a, 1998b), and a U deposit in Australia (Short et al., 1998) were attributed to association of the contaminants with a mobile colloidal phase. McDowell-Boyer et al. (1986) reviewed particle transport in porous media and pointed out that in addition to its importance in sanitary and geological engineering applications, groundwater colloids appeared capable of enhancing transport of contaminants that had a high affinity for sorbing to aquifer solids. A review by McCarthy and Zachara (1989) focused attention on colloid-
facilitated contaminant transport. Subsequent reviews of the subject have highlighted the growing body of literature on colloid transport (Ryan and Elimelech, 1996; Kretzschmar et al., 1999). Much of the progress in understanding groundwater colloids can be attributed to coordinated and interdisciplinary research programs at the national and international level. For example, the Commission of European Countries’ program on performance assessment of nuclear waste repositories included a subgroup (dubbed the Co-Co Club), which focused on the role of colloids and complexants on radionuclide transport. This group organized the first systematic study of colloids in the subsurface. The Grimsel colloid exercise (Degueldre et al., 1989) examined colloids in a series of sites within crystalline rock in the Swiss Alps. The study mobilized a dozen research groups using multiple methods to characterize the nature and abundance of colloids (Fig. 1). They demonstrated the polydispersity of groundwater colloids, but found that the particle size distribution of colloids conformed to a power law. Other funding agencies, such as the USDOE’s Subsurface Science Program, also organized multidisciplinary groups to examine colloid transport in groundwater. Subsequently, a large number of national and international workshops and symposia have been organized to share new results and help promote cooperation among colloid researchers. The workshop in Tjele, Denmark that has resulted in this special section is the latest effort in this direction and has the distinction of focusing attention on colloid transport in the vadose zone.

COLLOID TYPES AND CHARACTERISTICS

A variety of inorganic and organic materials exist as colloids and small particles in groundwater, including mineral precipitates (notably iron, aluminum, calcium,
and manganese oxides, hydroxides, carbonates, silicates and phosphates, but also including oxides and hydroxides of actinide elements such as U, Np, and Am), rock and mineral fragments (including layer silicates, oxides, and other weathering products of mineral phases), biocolloids (including viruses, bacteria, and protozoans), microemulsions of nonaqueous phase liquids, and macromolecular components of natural organic matter (including some components of humic substances and other polymers such as exocellular biopolymeric material secreted by microorganisms). This paper will focus primarily on inorganic colloids and biocolloids.

Considering this diversity of materials, it is worth noting that much of our understanding of colloid transport is based on studies of model colloids, which are not necessarily representative of the complex shapes and surface characteristics of natural environmental colloids. For example, even in model systems of latex particles and glass beads, particle deposition has frequently been observed to be greater than that theoretically predicted, and surface roughness has been considered a possible cause (Tobiason, 1989; Elimelech and O’Melia, 1990). This association is consistent with theoretical and experimental results by Suresh and Walz (1996, 1997), who demonstrated that surface roughness on latex microspheres reduced electrostatic repulsion. In their model system, roughness of only tens of nanometers resulted in the lower repulsive barrier, suggesting that roughness could have a significant impact on transport of irregularly shaped natural colloids. Spatial heterogeneity in surface chemistry is another characteristic of natural systems, and the impact of patch-wise heterogeneity of surface properties of colloids transport will be discussed below.

Bacterial transport is of interest as both a pathogen risk and a remedial strategy to degrade or immobilize contaminants in groundwater. However, the bacterial cell wall is structurally and chemically more complex and heterogeneous than the surfaces of inorganic colloidal particles. The cell surface is highly dynamic, responding to a variety of environmental changes. For example, cell size, hydrophobicity, exopolymer production, and other factors that affect transport can change in response to starvation (Heise and Gust, 1999). From an electrostatic perspective, bacteria must be viewed as soft particles in which ions can penetrate through the thickness of the surface appendages on the cells, and thus require a fundamentally different description of surface interaction forces than for ion-impenetrable inorganic colloids (Poortinga et al., 2002). Exocellular polysaccharides can promote attachment by bridging between the bacteria and aquifer surface, or inhibit it by preventing the bacteria from getting close to the surface. The role of these forces will vary, depending on the aqueous chemistry of the groundwater (Rijnaarts et al., 1999). Furthermore, individual cells within a bacterial population express different surface properties (Simoni et al., 1998), and surface charges and attachment probabilities can vary even within a single monoclonal bacterial population (Dong, 2002).

**SAMPLING GROUNDWATER COLLOIDS**

As the role of colloids in contaminant migration began to be recognized, it became quickly obvious that meaningful and accurate descriptions of subsurface colloids required a reevaluation and modification of groundwater sampling techniques typically accepted at the time. Groundwater samples are routinely obtained by pumping from wells. However, drilling, well construction, and pumping can greatly disturb the groundwater system. Drilling redistributes materials, creates fine particles, and provides a conduit for air to contact groundwater. Interaction between groundwater and the well screen, sand-pack, or casing can cause chemical changes. Sampling procedures can introduce artifacts, including the shearing of attached colloids at high pumping rates (Puls and Powell, 1992) and creation of colloids by alteration in chemical and physical conditions, such as alterations of O₂ and CO₂ content, temperature, pH, and redox potential. For example, introduction of O₂ into Fe(II)-rich groundwater can lead to production of Fe(III) oxides colloids (Liang et al., 1993).

New protocols were formulated to address these concerns (Backhus et al., 1993; McCarthy and Degueldre, 1993). Gentle sampling at low purging and pumping rates was advocated to minimize shear forces at or near the well screen. The criteria for selecting pumping rates for a given well was to minimize draw-down of the water elevation in the well, so that the pumping rate was matched to the rate of natural recharge into the well. Furthermore, rather than purging a fixed number of well volumes to remove stagnant water in the well casing, the criteria for adequate purging became stabilization of water quality parameters, such as pH, O₂, specific conductance, redox potential, and turbidity. Exclusion of atmospheric O₂ during sampling is also important for anaerobic samples.

While these precautions have been broadly implemented for sampling of colloids in saturated groundwater systems, the concept of a representative colloidal sample is still problematic because the nature and abundance of colloids can be influenced by seasonal or storm-related variations in flow and chemistry, especially in unconfined aquifers or in fast-flow karst systems. Sampling and analysis of microorganisms in groundwater are especially problematic because traditional analytical methods (APHA, 1995) are based on culturing of viable cells and many pathogens of interest are not readily cultureable in the laboratory (Maier et al., 2000). This problem has been partly alleviated by the development of nucleic acid-based analyses, which can detect both viable and nonviable microorganisms (Abbadaszadegan et al., 1993, 1999a; Josephson et al., 1993).

The situation for colloid sampling in the vadose zone is even more uncertain because there have been few, if any, comparable studies on appropriate and effective techniques for sampling colloids in the vadose zone. This remains a critical challenge.
COLLOID TRANSPORT IN SATURATED SYSTEMS: THE PROBLEM OF HETEROGENEITY

While a theoretical framework exists for describing colloid formation, stability, and transport in model systems, the principal scientific issue that limits prediction is understanding of how colloids behave in natural subsurface systems. Natural systems have complex solution chemistry and include mixed colloidal phases (e.g., layer silicates, Fe-, Al-, and Si-(hydr)oxides and natural organic matter). The kinetics of colloid formation and dissolution, and the crystallinity, surface characteristics, and colloidal stability of mineral oxides are affected by the presence of specific solutes and other colloids. Furthermore, chemical and physical heterogeneity affects colloid behavior at a range of spatial and temporal scales. At the microscopic scale, detachment and mobilization rates are affected by factors such as the arrangement and nature of surface functional groups, surface hydrophobicity, and surface roughness, as well as by the physical arrangement of the colloidal phases relative to the larger grains and pore spaces. Colloid transport at the field scale will be affected by the spatial distribution of physical and chemical features along a flow path in an aquifer. Some examples of these issues are discussed below.

Heterogeneity in Surface Charge

Filtration theory (Yao et al., 1971) has proven effective in predicting colloid attachment under conditions where colloid–grain interactions are attractive, but subsurface environments most often involve conditions unfavorable to colloid deposition due to electrostatic repulsion between the generally negatively charged media and colloids. The extent of colloid transport is generally described in terms of the magnitude of the surface charge on the colloids and aquifer surfaces. However, spatial heterogeneity in surface charge can render the average surface charge of the grains (as measured by the zeta ($\zeta$) potential) irrelevant to transport predictions. This was very effectively illustrated by Elimelech et al. (2000), who created patchwise heterogeneity by treating quartz grains with aminosilane, which reversed the surface charge from negative to positive. Mixing the aminosilane-modified and unaltered quartz grains in varying proportions created chemically heterogeneous porous media. The measured $\zeta$ potential increased with the increasing abundance of aminosilane-modified grains in the media (Fig. 2a). Conversely, the breakthrough of negatively charged 0.3-μm silica colloids was reduced as the proportion of the positively charged aminosilane-modified grains increased (Fig. 2b). However, these results differed substantially from those that would have been predicted based on the average $\zeta$ potential of the media. Theoretical collision efficiencies calculated for the measured $\zeta$ potential of porous media containing 0 to 50% aminosilane-modified grains predicted essentially complete breakthrough of the colloids ($C/C_0 = 1$). At the pH of the experiments, the $\zeta$ potential of the media containing 75 or 100% of the aminosilane-modified grains was positive, so all of the colloids should have been retained ($C/C_0 = 0$). Thus, the theoretical prediction of a dichotomous result of either complete breakthrough or complete retention is clearly at odds with the observed breakthrough curves (Fig. 2b). Colloid retention was controlled by attachment of individual aminosilane-modified grains, regardless of the average surface properties of the media. The results demonstrated the insignificance of the average $\zeta$ potential of chemically heterogeneous porous media. Although the extent of patchwise heterogeneity to colloid transport is clearly established in this study, characterization of...
the chemical heterogeneity is difficult even in laboratory-scale studies, and impractical in field studies.

**Colloid Transport and Aqueous Chemistry**

The importance of ionic strength and composition on colloid transport is well established. The expected effect—increased deposition at higher ionic strengths, and much greater retention in the presence of divalent, compared with monovalent, cations of comparable ionic strengths—is attributed to screening of repulsive surface interaction energies between colloids and grain surfaces. These relationships have been extensively demonstrated in laboratory studies, including in experiments with natural soil material (Grolimund et al., 1998). Particle deposition rates with solutions of Ca\(^{2+}\) were shown to be higher than with Na\(^{+}\), and the critical deposition concentrations with Na\(^{+}\) were greater than those with Ca\(^{2+}\). Although Grolimund et al. (1998) used natural soil, the columns were constructed using repacked, 2-mm soil particles. The aqueous chemistry within the column was controlled during the experiments, a precaution that was facilitated by the noncalcareous nature of the soil and its low cation exchange capacity (CEC).

While aqueous chemistry may control the interactions between colloids and geological media, chemical and physical properties of natural geological materials may in some cases control the aqueous chemistry, and thus the extent of colloid transport. The role of pore structure and geochemical processes was illustrated in a study of microsphere transport in intact monoliths of saprolite, which is a highly weathered, decomposed material that retains the fabric of the parent rock (McCarthy et al., 2002). As was found by Grolimund et al. (1998), recovery of the microspheres consistently decreased with ionic strength, but a much greater concentration of Na\(^{+}\) in the influent solution was required to result in a similar reduction in colloid recovery as compared with Ca\(^{2+}\) (Fig. 3a).

Interpretation of those data was complicated, however, because the composition of the effluent solution was significantly altered along the flow path through the monolith (Fig. 3b). The change in aqueous chemistry was attributed in part to the significant CEC of the saprolite (140 mmol kg\(^{-1}\)), although the inputs of the monovalent influent solution before the introduction of the colloid injections greatly exceeded the total 1800 mmol of charge in the monolith. Of greater importance was the influence of the pore structure on the kinetics of solute interactions. While the total porosity of the saprolite was 49%, advective flow was primarily through the fractures, which constituted only about 5% of the porosity. Thus, failure to achieve a monocationic system during experiments with Na\(^{+}\) input solutions (Fig. 3b) was attributed to slow diffusive exchange between the pore water in the fractures and in the fine-grained, Ca- and Mg-rich matrix. Divalent cations were released at concentrations sufficiently high that surface potentials would be determined solely by the divalent cation concentration, even during inputs of the Na solutions. The dominance of the divalent cations in colloid transport is

![Fig. 3. Transport of four sizes of colloids through a saprolite monolith using input solutions of varying cation charge and ionic strengths.](image-url)
illustrated by the similarity in the slopes of the colloid stability curves (Fig. 3c). The greater recoveries of colloids in the Na experiments is likely related to the creation of a gradient of Na- to Ca-dominated chemistry along the flow path, resulting in shorter effective travel distances for colloids with a Ca-dominated solution chemistry.

**Colloid Transport in Fractures and Karst**

In addition to the effects described above of pore-scale exchange on colloid transport within dual-porosity systems, larger-scale structural heterogeneities, such as fractured or karstic formations, can function as superhighways for rapid transport of colloids. Karst aquifers, which typically occur in carbonate rocks, contain solution-enlarged fractures or conduits and often experience very rapid flow (kilometers per day) during and after heavy rains (White, 1988). Under these conditions large amounts of colloids and sediments can be suspended and rapidly transported. Studies of spring discharge in the marble karst of the Colonnata basin in Italy indicate annual sediment yields in the order of $100 \text{ t km}^{-2}$ (Drysdale et al., 2001). Karst systems can transport a wide range of particle sizes and types (Atteia and Kozel, 1997), which can include pathogenic microorganisms as well as contaminants such as metals, that are attached to the suspended particles (Vesper and White, 2003). These systems are highly dynamic, with changes in suspended load a function of both physical factors such as flow rate and conduit size and chemical factors such as the pH of the infiltrating water (Shevenell and McCarthy, 2002). Understanding these processes and making even the most simple predictions of their behavior is still a major challenge.

Rapid transport of colloids can also occur in fractured materials, even those that are typically considered aquitards, rather than aquifers. This was demonstrated by a series of solute and colloid field tracer experiments that were performed in fractured, fine-grained glacial tills in Canada (McKay et al., 1993) and Denmark (McKay et al., 1999), and in a highly weathered shale saprolite in Tennessee (McKay et al., 2000). All of the materials had low to moderate hydraulic conductivity values (geometric mean of approximately $10^{-7} \text{ m s}^{-1}$) and flow was dominated by the fracture systems (Fig. 4). Solute and colloid tracers were introduced under conditions simulating natural flow. First arrival of the colloidal tracers (latex microspheres and bacteriophage) in the samplers occurred at rates of 2 to $5 \text{ m d}^{-1}$ over a distance of 4 m in the Canadian till, and 1 to $6 \text{ m d}^{-1}$ over a distance of 1.4 to 3.4 m in the Danish till. In the saprolite experiment, first arrival rates for the colloidal tracers ranged from 5 to $200 \text{ m d}^{-1}$, over distances of 2 to 35 m. In all cases the colloidal tracers arrived much earlier than the nonreactive solute tracers (bromide, chloride, or fluorescent dye), typically by factors of 10 to 100. This was largely due to diffusion of the solutes, but not the colloids, into the fine pore structure. These experiments demonstrate that colloid transport can occur very rapidly in a wide range of types of fractured materials and suggests the need for research on colloid entrainment and retention processes under these conditions.

**COLLOID TRANSPORT IN THE VADOSE ZONE**

Understanding colloid transport in partially saturated porous media is a major challenge that is only beginning to be addressed. In addition to all the difficulties posed by chemical and physical heterogeneity in water-saturated systems, predictions of colloid and pathogen transport in the vadose zone are further complicated by the presence of an air phase in addition to the solid and water phases present in saturated media. Challenges exist at a range of spatial scales, from the microscopic (interactions of colloids with the air-, water-, and solid interfaces) to the macroscopic (preferential flow and fingering), and include kinetic considerations on the resupply of colloids following transient infiltration events. In spite of the relevance of colloids to many environmental problems in the vadose zone, there has not been a strong focus on this topic by research groups or funding agencies. For example, colloids were barely mentioned...

Many of these issues are addressed in papers of this special issue, and will be addressed only briefly to place the results of specific studies in broader perspective.

**Colloids and the Air, Water, Solid Interfaces**

The interaction of colloids with the air–water (AW) interface has been invoked as a dominant process in colloid retention (Wan and Wilson, 1994a). Wan and Tokunaga (1997) also proposed an additional colloid retention mechanism, film straining, which hypothesizes that transport of suspended colloids can be retarded due to physical restrictions imposed when the thickness of water films falls below the diameter of the colloids.

Most subsequent studies of colloid transport under conditions of partial saturation have been based on mass balance of colloid breakthrough in packed sand columns (Wan and Wilson, 1994b; Schaffer et al., 1998; Jewett et al., 1999; Jin et al., 2000; Lenhart and Saiers, 2002), rather than direct visualization. Reduced colloid transport at lower water contents (and presumably greater areas of AW interfaces) was attributed to sorption at the AW interface and film straining (Corapcioglu and Choi, 1996; Lenhart and Saiers 2002). However, recent studies suggest that the assumed dominant roles of the AW interface and film straining should be reevaluated.

It has been generally assumed, based on visualization studies of Wan and Wilson (1994a), that retention at the AW interface was relevant to a wide variety of colloids, including hydrophilic and hydrophobic latex microspheres, clay particles, and bacteria and viruses (Wan and Wilson, 1994b; Corapcioglu and Choi, 1996; Schaffer et al., 1998; Lenhart and Saiers, 2002; Gamedinger and Kaplan, 2001; Sirivithayapakorn and Keller, 2003). More recently, Wan and Tokunaga (2002) demonstrated in bubble column experiments that only positively charged particles attached to the negatively charged AW interface, suggesting that immobilization at the AW interface would be limited to a smaller subset of environmental colloids. Wan and Tokunaga (1997) conceptualized film straining as a colloid retention process within thin films of water flowing from one pendular ring to the next. However, Lenhart and Saiers (2002), noting that calculated water film thicknesses were 20-fold smaller than their 0.4-μm colloids, concluded that the relevant immobilization process was the degree of pendular ring discontinuity, rather than immobilization in thin water film.

Crist et al. (2004) reported a real-time, pore-scale, three-dimensional visualization technique to examine colloid movement through a packed sand column. They observed that hydrophilic, negatively charged latex colloids were not retained at either the AW or the solid–water (SW) interfaces. Instead, the colloids concentrated at the edge of pendular rings, that is, within but not attached to the thin film of water between the AW interface and the surface of the sand grain. Although the importance of distinguishing between colloid attach-

**Kinetic Constraints on Colloid Mobilization**

Water movement in natural soils is dominated by transient infiltration events rather than steady-state unsaturated flow. Infiltration of water through intact macroporous soil was shown to result in mobilization and transport of previously deposited, in situ colloids (Jacobson et al., 1997; de Jonge et al., 1998; Schelde et al., 2002). Furthermore, subsequent irrigation events could leach more colloids, although the amount of colloids mobilized in the subsequent wetting was dependent on the length of the flow delay. The amount of colloids leached appeared unrelated to the irrigation rate, suggesting that hydrodynamic shear was not involved.

The data demonstrate that irrigation can result in rapid mobilization and transport of an initial peak of colloids. As infiltration continues, the supply of readily mobilized colloids in the soil macropores is depleted. Between irrigation events, a time-dependent process replenishes the supply of readily mobilized colloids. Schelde et al. (2002) developed an equivalent macropore model that hypothesized that the resupply of colloids was a diffusion-limited process governed by two processes, one involving time-dependent release of colloids from the soil matrix to the crust at the macropore wall, followed by diffusion of those colloids through a layer of immobile water adjacent to the crust. The magnitudes of the mass transfer coefficients reported in these studies raise some question as to true origin of the time-dependent process(es). However, the work clearly highlights the need for additional experimental and theoretical studies to resolve the mechanistic basis for the time-dependent resupply of mobile colloids. Kjaergaard et al. (2004) examined the effect of clay content and soil water potential on colloid mobilization. Their results make clear that understanding of colloid mobilization in natural structured soil will be a complex challenge that must encompass both intrinsic soil properties and dynamic processes controlling soil aggregate structure and stability, macropore development, and colloid dispersion.

**Episodic Flow, Preferential Flow, and Fingering**

Most current models of water, solute, and colloid movement in partially saturated media are based on the traditional Richards equation without hysteresis in the soil water characteristic curve, which is based on uniform flow in a continuum (Milly, 1988). However, infil-
Moisture content visualization of finger formation and persistence in a sand slab (Glass et al., 1989).

Infiltration at the soil surface is often highly episodic, which can lead to large changes in the water table elevation or even development of perched water tables (i.e., transient saturated conditions underlain by unsaturated materials). This is especially common in fractured or macropore-dominated materials, where filling and drainage occur much more rapidly in the fractures and macropores than in the fine-grained soil matrix (NRC, 2001). As a result, flow rates and shear stresses, which can mobilize colloids, tend to vary over a much wider range in the vadose zone than in the saturated zone.

Rapid infiltration, especially in materials that initially have a relatively low degree of saturation, can also induce flow instabilities or fingers (Fig. 5). Development of fingering can result in rapid downward flow and the efficient transport of colloids through the vadose zone, even under bulk moisture conditions where traditional theory (uniform flow) would predict very slow migration. Also, when the bulk moisture content (outside the fingered flow paths) is below the critical moisture content for interconnected water films, much thicker water films along grain surfaces or more hydraulically conductive fingers (DiCarlo et al., 1999) are postulated to exist within the flow paths.

Layered structural interfaces can also initiate development of fingered flow paths by creating capillary fringes when a fine layer overlies the coarse-grained layer (Kung, 1990a, 1990b; Steenhuis et al., 1991). Likewise, macropores and structural cracks will transport water to deeper depths with locally high moisture contents (Andreini and Steenhuis, 1990; Flury 1996). Episodic infiltration, preferential flow, and transient saturation at fingertips are important topics of research that could significantly alter predictions of the rate and extent of colloid transport through the vadose zone.

Microbial Transport

Understanding the transport behavior of microorganisms is particularly important in the vadose zone because the majority of pathogens found in drinking water supplies likely originated from sources such as septic fields, land application of sewage sludge or animal waste, or leaking sewer lines. Although there is a substantial body of published research on microbial transport in the subsurface, much of it focuses on transport in the saturated zone (e.g., Keswick et al., 1982; Bales et al., 1989; Fontes et al., 1991; Hornberger et al., 1992; McKay et al., 1993; Pieper et al., 1997; Harvey, 1997; Chu et al., 2000; Ginn et al., 2002), rather than on transport in the vadose zone (e.g., Lance and Gerba, 1984; Powelson et al., 1990; Schafer et al., 1998; Jin et al., 2000; Chu et al., 2001). The vast majority of the published studies are based on laboratory experiments or modeling studies, with very few based on actual field investigations of microbial transport in the vadose zone.

In addition to the uncertainties concerning scaling of findings from the lab scale to the field scale, there are a number of other basic research issues relevant to both the saturated and vadose zones that still need to be addressed. One such issue is the influence of microbial shape, size, and surface characteristics on transport. Microbial surfaces are often very complex, with cell walls ranging from loosely held slime layers to rigid capsules, frequently with protrusions such as spikes, fimbria, and flagella (Maier et al., 2000). The protrusions can play an especially important role in attachment to soil particles. Poortinga et al. (2002) argued that this is one of the reasons that DLVO-based descriptions often fail to describe microbial adhesion. Even within monoclonal populations, the size and surface characteristics of bacteria can vary widely, affecting transport behavior (Heise and Gust, 1999; Simoni et al., 1998; Dong, 2002). Another problem is that there is very little information on the actual occurrence of pathogens in soils or groundwater.

Instead, most environmental monitoring of microorganisms is based on indicator bacteria, such as Escherichia coli and fecal coliforms (USEPA, 1986; Maier et al., 2000). Not only are many strains of these indicator organisms...
nonpathogenic to humans, they may not be good indicators for the occurrence of other types of pathogens, such as protozoa and viruses. Only recently have there been any studies of the actual occurrence of pathogenic viruses in groundwater (Abbadaszadegan et al., 1999b; Azadpour-Keeley et al., 2003). Research on microbial transport also tends to focus on a just a few types of nonhazardous microbial or nonmicrobial tracers, such as bacteriophage PRD1 and MS2, E. coli, Pseudomonas syringae, indigenous bacterial strains, and various yeast cells or fluorescent microspheres (Keswick et al., 1982; Harvey, 1997). As a result, in many cases it isn’t clear how widely applicable the results of these studies are to transport of actual microorganisms of environmental interest, such as specific pathogenic strains, or bacteria added to soils to increase contaminant degradation.

Other issues that need to be addressed include transport and survival of pathogens at large spatial and temporal scales. For example, performance of wastewater infiltration systems or land application of sewage sludge is likely to be influenced by the physical scale of the operation, the magnitude of water fluxes through the system, and the duration of operation. Intensive application of wastes has the potential to substantially change the physical, chemical, and biological properties of the vadose zone, and hence change its ability to filter out pathogens.

**SUMMARY AND CONCLUSIONS**

This introductory paper has presented a general, and admittedly subjective, perspective on some of the key challenges in colloid transport that have been, or need to be, addressed. These are summarized in Tables 1 and 2.

Two general topics were highlighted. First, the problem of chemical and physical heterogeneity in natural systems remains a key challenge in describing colloid transport and retention in saturated groundwater systems. Although the effect of individual factors, such as the surface charge of porous media, solution chemistry, or flow regime, can be fairly well described, these factors are seldom uniform andunchanging across spatial and temporal scales that are relevant to environmental processes. Practical methods do not currently exist for characterizing the distribution of those factors in natural systems, making it difficult to develop robust descriptions of behavior in natural systems. However, even given these limitations, it is important to recognize that research in complex systems is needed to at least help us to understand what we do not understand. For example, insights into the irrelevance of $\xi$ potential in predicting colloid behavior in patchwise heterogeneous media (Elimelech et al., 2000) is critical to avoiding fallacious predictions. Likewise, it is important to realize that geochemical and pore-scale structural properties of a formation can have complex, and sometimes nonintuitive effects on colloid behavior. Slow diffusive exchange processes can strongly buffer against any effort to manipulate colloid behavior through imposed changes in solution chemistry (McCarthy et al., 2002). However, exchange of meteoric water and solutes in fine pores during storm events can also create rapid changes in solution chemistry that can result in short-term pulses of colloid mobilization and transport (Shevenell and McCarthy, 2002). New modeling approaches may be successful in addressing some of these problems. For example, inverse modeling theory may help identify the key processes controlling colloid transport in heterogeneous porous media. The technique has been used extensively in groundwater flow and solute transport. Comprehensive reviews have been provided by Yeh (1986), Kuiper (1986), Carrera (1998), and Cushman (1990). Marley and Townley (1996). Using colloid arrival breakthrough curves, a forward colloid transport model can be used to solve a generalized inverse problem by identifying and prioritizing the subset of potential geologic and geochemical parameters and parameter structure needed to construct an accurate realization of colloid transport, and thus determine the most probable parameters and processes that are needed to accurately predict colloid transport (Sun, 1999; Sun et al., 2001; Tsai et al., 2003).

The second category of challenges, and the one that is the focus of this special issue, is colloid transport in the vadose zone. In many ways, it is fair to say that, with respect to the vadose zone, we don’t really even know all the questions, let alone understand the an-

---

**Table 1. Key challenges in saturated groundwater.**

<table>
<thead>
<tr>
<th>Issue</th>
<th>Implementation</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobilization of colloids during sampling</td>
<td>Broad application</td>
<td>Well recognized</td>
</tr>
<tr>
<td>Patch-wise Heterogeneity</td>
<td>Surface properties</td>
<td>Some success</td>
</tr>
<tr>
<td>Electrostatics</td>
<td>Lab and models, Field application, Solution chemistry</td>
<td>Limited, qualitative</td>
</tr>
<tr>
<td>Geochemical or structural</td>
<td>Lab and field</td>
<td>Generally ignored</td>
</tr>
<tr>
<td>Fracture flow</td>
<td>Structural heterogeneity, Intact monoliths, field scale</td>
<td>Descriptive at best</td>
</tr>
</tbody>
</table>

---

**Table 2. Key challenges in the vadose zone.**

<table>
<thead>
<tr>
<th>Issue</th>
<th>Implementation</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distinguishing entrained- and retained colloids</td>
<td>Limited</td>
<td>Generally ignored</td>
</tr>
<tr>
<td>Temporal variations</td>
<td>Not well understood</td>
<td>Generally ignored</td>
</tr>
<tr>
<td>Air-water interface</td>
<td>Micromodels, columns</td>
<td>Accepted mechanism</td>
</tr>
<tr>
<td>Air-water-solid interface</td>
<td>New concept</td>
<td>Is it real? Implications?</td>
</tr>
<tr>
<td>Diffusive limitations</td>
<td>Macroporous flow</td>
<td>Needs broader investigation</td>
</tr>
<tr>
<td>Preferential flow</td>
<td>Heterogeneous and structured media</td>
<td>Unexplored for colloid transport</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

We thank R.W. Gentry for reviewing the manuscript and valuable discussion, and the Center for Environmental Biotechnology of the University of Tennessee for partial support of this effort. We also thank M. Fluory and two anonymous reviewers for comments that improved the manuscript.

REFERENCES


