Going even smaller: Engineering sub-5 nm nanoparticles for improved delivery, biocompatibility, and functionality

Manman Xie, Emory University
Yaolin Xu, Emory University
Jing Huang, Harvard Medical School
Yuancheng Li, Emory University
Liya Wang, Emory University
Lily Yang, Emory University
Hui Mao, Emory University

Journal Title: WILEY INTERDISCIPLINARY REVIEWS-NANOMEDICINE AND NANOBIO TECHNOLOGY
Volume: Volume 12, Number 6
Publisher: WILEY | 2020-05-20, Pages e1644-e1644
Type of Work: Article | Post-print: After Peer Review
Publisher DOI: 10.1002/wnan.1644
Permanent URL: https://pid.emory.edu/ark:/25593/vts11

Final published version: http://dx.doi.org/10.1002/wnan.1644
Accessed September 26, 2023 4:32 AM EDT
Going Even Smaller: Engineering Sub-5 nm Nanoparticles for Improved Delivery, Biocompatibility and Functionality

Manman Xie¹, Yaolin Xu¹, Jing Huang², Yuancheng Li¹, Liya Wang¹,³, Lily Yang⁴, Hui Mao¹,*

¹Department of Radiology and Imaging Sciences, Emory University, Atlanta, Georgia, The United States of America
²Laboratory of Vascular Biology, Harvard Medical School, Boston, Massachusetts, The United States of America
³Department of Radiology, The People’s Hospital of Longhua, Shenzhen, Guangdong, China
⁴Department of Surgery, Emory University, Atlanta, Georgia, The United States of America

Abstract

The rapid development and advances in nanomaterials and nanotechnology in past two decades have made profound impact in our approaches to individualized disease diagnosis and treatment. Nanomaterials, mostly in the range of 10–200 nm, developed for biomedical applications provide a wide range of platforms for building and engineering functionalized structures, devices, or systems to fulfill the specific diagnostic and therapeutic needs. Driven by achieving the ultimate goal of clinical translation, sub-5 nm nano-constructs, in particular inorganic nanoparticles such as gold, silver, silica and iron oxide, have been developed in recent years to improve the biocompatibility, delivery and pharmacokinetics of imaging probes and drug delivery systems as well as in vivo theranostic applications. The emerging studies have provided new findings that demonstrated the unique size-dependent physical properties, physiological behaviors and biological functions of the nanomaterials in the range of the sub-5 nm scale, including renal clearance, novel imaging contrast and tissue distribution. This advanced review attempts to introduce the new strategies of rational design of engineered nanoparticles with the core sizes under 5 nm with consideration of the clinical and translational requirements. We will provide readers the update on recent discoveries of chemical, physical and biological properties of some biocompatible sub-5 nm nanomaterials as well as their demonstrated imaging and theranostic applications, followed by sharing our perspectives on the future development of this class of nanomaterials.

Graphical Abstract

*Corresponding Author: Hui Mao, PhD, ¹Department of Radiology and Imaging Sciences, Emory University, Atlanta, Georgia 30329, USA, hmiao@emory.edu.

Competing Interests
The authors have declared that no competing interest exists.
1. Introduction

The rapid development and advances in nanomaterials and nanotechnology in past two decades have made the significant impact in the field of biomedical research [1–2]. Nanomedicine not only has changed our approaches towards individualized disease diagnosis and treatment, but also provided promising new solutions to various challenging medical problems [3–6]. Nanomaterials developed for biomedical applications enable us to design and build a wide range of functionalized nanostructures, devices, or platforms for innovative diagnostic and therapeutic as well as integrated theranostic applications [7–8]. For in vivo applications, such as imaging and drug delivery, nanomaterials, which are mostly in the range of 10–200 nm, provide unique properties and advantages, including: 1) enabling micro-scale or surface chemistry that combines/assembles multiple functions or applications on a single platform; 2) delivering large payload of imaging or therapeutic agents; 3) prolonging blood circulation or altering pharmacokinetics and biodistribution for improved delivery; and 4) introducing target specificity to a molecular biomarker or diseased tissue. These desirable features can be achieved through engineering nanomaterials with rational design and synthesis for specific shape, size, compositions and surface, which are “synthetic...
identity” controlled at the preparation of the materials [9–10]. While a significant amount of early studies focused on developing nanomaterials with superb chemical and physical properties for high performance and enhanced functions, there has been increasing effort on better understanding biological and physiological properties or “biological identity” of nanomaterials, which largely dictate the fate of nanoparticles and their behaviors in the biological environment, no matter intended or unwanted, once they enter the living systems and interact with the biological environments [11–13].

Extensive investigations have shown that the size of nanoparticles not only plays important roles in the physical and chemical, but more importantly also in the biological properties of the nanomaterials [10, 14–17]. For example, the size of nanoparticles strongly affects the surface area and a number of functional groups for conjugating ligands to target specific biomarkers or cells, subsequently changing the surface charges that can alter blood retention time, biodistribution and cellular internalization [15–19]. Worth noting, the biological and physiological properties of nanomaterials are not only size-dependent but often non-linearly dependent on different hosting biological systems [20], such as cells, vessels, tissues, organs and vasculature. Unlike most physical and chemical properties of nanoparticles that are positively correlated to the size, i.e., stronger or higher performance with increased sizes, the biological properties and behaviors that are important to the in vivo imaging and theranostic applications are often reversely correlated to the size of the nanoparticles. Typically, nanoparticles with smaller size are more biocompatible and favorable to the living systems, causing less perturbation to the normal biological and physiological functions. Larger nanoparticle size contributes greatly to higher accumulation in the organs of the reticuloendothelial system (RES) [17], shortened blood half time [18], poorer delivery [19] and slower degradation and clearance [19] when nanoparticle imaging probes and drug carriers are administered systemically. Preferential uptake of larger nanoparticles by immune cells, such as macrophages, may lead to more pronounced nanoparticle-induced immunogenicity and unwanted side effects to the patients [21]. Poor delivery and the side effects due to slow clearance of engineered nanomaterials have now becoming major concerns for further development of in vivo applications of nanomedicine. Therefore, there has been an increasing effort in seeking solutions to overcome obstacles for clinical translations of nanoparticle imaging probes and theranostic platforms, particularly inorganic nanoparticles.

Going even smaller than the typical core size range of 10–200 nm, sub-5 nm nano-constructs, which were termed as ultrasmall, extremely small, exceedingly small or ultrafine by different studies, have been developed in recent years to improve the biocompatibility, delivery and pharmacokinetics of nanoparticles [22–26]. The new findings from earlier studies demonstrated the unique size-dependent physical properties, physiological behaviors and biological functions of the various types of inorganic nanoparticles, such as gold, silver, gadolinium and iron oxide nanoparticles, in the range of the sub-5 nm scale [27–34], including renal clearance, novel imaging contrast and tissue distribution. Here, we will provide an overview and update of the recent development and discoveries of chemical, physical and biological properties as well as applications of some biocompatible imaging and theranostic nanomaterials with the core sizes under 5 nm. The perspective for further
developing this class of nanomaterials and their potential clinical translation will also be discussed in this review.

2. Challenges for Making Water-Dispensable and Stable Sub-5 nm Nanoparticles

2.1. Size Uniformed Sub-5 nm Crystalline

At the sub-5 nm extremely small size scale, controlling the size and crystallinity requires more precise chemistry and experimental conditions. In the early years of development and investigation to understand fundamentals in nanomaterials, sub-5 nm nanoparticles were often prepared as the clusters of atoms/molecules using noble metals [35–38], such as Au, Ag, Pd, assembled on the solid phase matrices. Among those, Au nanoparticles have been studied the most extensively to understand the mechanisms and approaches for precisely controlling the cluster sizes and inherited size-dependent fluorescent and plasmonic properties, leading the way for developing other types of sub-5 nm nanocomposites for specific biomedical applications. With increased knowledge and continuous efforts in advancing nanochemistry, nowadays a variety of sub 5 nm nanoparticles, with examples shown in Table 1 [39–53], can be made through methods refined from the traditional approaches that have been widely used for making larger nanoparticles, either through thermal composition or decomposition at specific temperature and organic solvents or co-precipitation in the aqueous media. The former method has advantages of producing nanoparticles with high crystallinity and size uniformity, while the latter uses mild condition for generating nanoparticles with a range of colloidal forms to allow for selecting nanocrystals with specific sizes.

For example, extremely-small-sized iron oxide nanoparticles (ESIONs) with core sizes of 1.5, 2.2 and 3 nm, as shown in Figure 1A, can be prepared via the thermal decomposition of iron–oleate complex in the presence of oleyl alcohol [25, 41]. Using oleyl alcohol enables the reduction of iron–oleate complex under relatively low reaction temperature, resulting in the production of ESIONs of 2 or 3 nm core with a maghemite crystal structure as revealed by the XRD pattern (Figure 1B). The ESIONs prepared in the organic solvent can be then transferred to the water after capping with the poly(ethylene glycol)-derivatized phosphine oxide (PO-PEG) ligands to make the particles water-dispersible. The magnetization of ESIONs was found decreased as the core size decreases (Figure 1C), leading to emerging bright contrast enhancement in $T_1$-weighted magnetic resonance imaging (MRI). The thermal decomposition method can also be applied to synthesize other inorganic nanocomposites, such as Ag$_2$S [45] for CT imaging and Gd$_2$O$_3$ [42] for MRI. In general, types of precursors, reaction temperature and time and solvent conditions (e.g., pH and polarity) are considered to be the most critical factors in synthesis of sub-5 nm nanoparticles with controlled sizes.

Alternatively, Li et al. developed a high-temperature co-precipitation route to make monodispersed water-soluble and biocompatible ultrasmall magnetic iron oxide nanoparticles (UMIONs, core diameter = 3.3 ± 0.5 nm) for dual MRI contrast agents [54], namely brightening or positive $T_1$-weighted contrast and darkening or negative $T_2$-weighted
contrast in a single agent. Comparing to the thermal decomposition of organometallic precursors in organic solvents, the protocol modified from the traditional method makes co-precipitation take place in the selected coating polymer solution mixing with FeCl$_3$·6H$_2$O and FeSO$_4$·7H$_2$O dissolved in concentrated HCl and the iron precursor at high temperature before adding the precipitating agents, such as ammonia solution. These polymer ligands can effectively prevent the aggregation of formed nano-colloidal droplets to ensure a small particle size. The low molecular weight polymers, such as thiol functionalized poly(methacrylic acid) (PMAA-PTTM, $M_w = 7524$ g mol$^{-1}$), were used as a stabilizer to control particle size and size distribution. The resulting iron oxide nanoparticles were concentrated and then dialyzed to remove impurities, before collecting dried UMIONs.

### 2.2. Effective Surface Coating

One of the essential requirements for nanoparticles developed for *in vivo* applications is the stability in aqueous media at physiological conditions. Since most sub-5 nm nanoparticles are prepared in the non-aqueous phase to ensure the ultrafine size and highly uniformlyed crystallinity, they need to be transferred and stabilized in water. However, the surface area and surface energy of sub-5 nm nanoparticles are significantly higher than those of larger sized nanoparticles, leading to much stronger tendency to aggregate and subsequently precipitate. Comparing to the common strategies and methods used to stabilize larger nanoparticles with conventional high molecular weight coating materials and thick polymers, it is much more challenging to have a coating material to stabilize the nanoparticle cores while avoiding a significant increased overall size, which may affect the biodistribution and clearance of nanoparticles. Therefore, low molecular weight or small molecules, such as glucose, glucuronic acid, glutathione and citrate, have been used to coat the nanoparticle surface, either directly or further modified or polymerized, rendering the water solubility and stability of sub-5 nm nanoparticles as well as their surface functions [54–58]. Zwitterionic molecule glutathione, which can bind to the nanoparticle surface through the bond between thiol groups and metal ions as shown in Figure 2A, is a commonly used coating molecule for stabilizing sub-5 nm nanoparticles [32–33, 44–48]. For example, Ag$_2$S nanoparticles with the core size of 3 nm developed for X-ray CT contrast enhancing applications were coated with glutathione. Resulting Ag$_2$S nanoparticles with hydrodynamic diameter of 4.9 ± 1.0 nm and ζ-potential of −11 ± 2 mV exhibited no size increase after incubation in phosphate buffered saline (PBS) or PBS with 10% fetal bovine serum at 37 °C for 24 h [44]. Similarly, small hydrophilic molecule D-glucuronic acid was used to coat the synthesized paramagnetic ultrasmall gadolinium oxide (Gd$_2$O$_3$) nanoparticles with core size of ~1 nm [55]. The carboxylic group of D-glucuronic acid binds to surface metal ions Gd$^{3+}$ to stabilize Gd$^{3+}$ and the surface of nanoparticles. Thermogravimetric analysis showed that 84.55% surface of the ultrasmall Gd$_2$O$_3$ nanoparticles were covered by D-glucuronic acids to provide sufficient stability.

On the other hand, Huang *et al.* developed the “*in situ* polymerization” coating method as illustrated in Figure 2B to enable glucose to form a uniformed thin layer of oligosaccharides directly on the surface of hydrophobic sub-5 nm ultrafine iron oxide nanoparticles (uIONPs) (Figure 2C). Oligosaccharide coating not only renders uIONPs with excellent water solubility and high stability but also can be further modified through reaction with ammonia...
to introduce – NH$_2$ functional groups that can be used to cross-link or conjugate affinity ligands to target specific cells and biomarkers [27].

With proper selection and optimization, low molecular weight polymers containing functional groups can also be used to transfer and stabilize sub-5 nm nanoparticles in the water. For example, oleic acid (OA)-coated NaGdF$_4$ nanoparticles with the core size of 1.2 or 2.4 nm can be synthesized with a co-precipitation method [53]. Poly(acrylic acid) polymers (PAA, M$_W$ ≈ 1200, 2000, 5000, and 8000) were then capped on the particle surface with a two-step ligand-exchange method, first removing original OA molecules by treatment with (CH$_3$)$_3$OBF$_4$ in dimethylformamide, and then adding PAA molecules to cap the surface. The resulting PAA-capped GdOF nanoparticles (2.1 ± 0.2 nm), namely Gd-dots, are sufficiently stable to be used as MRI contrast agents with a remarkable ionic relaxivity of ~75 mM$^{-1}$ s$^{-1}$ in albumin solution at 0.5 T for magnetic resonance renography and angiography as tested in rabbits. These Gd-dots also exhibited efficient renal clearance with <3% of injected amount left 12 hours post-injection.

3. Improved Biocompatibility and Biodistribution

3.1. Reduce RES Uptake and Accelerate Clearance

The use of circulatory system for delivery of imaging probes and drugs is the most common approach to achieve proper distribution and retention of the systemically administered imaging and theranostics agents. For in vivo applications, engineered nanoparticles given to patients, especially inorganic nanoparticles, are expected to be completely cleared from the body within a reasonable time frame to avoid long-time toxicity [59–60]. Indeed, this is the essential requirement for getting approval from U.S. Food and Drug Administration (FDA). The intravenously (i.v.) injected nanoparticles are known to be cleared mainly through two routes: 1) taken up and then degraded in the RES organs (e.g., liver and spleen), and 2) secreted by the kidney. The two clearance mechanisms involve in different physiological barriers [60–64], thus leading to variations in the retention time in the body. Among many factors involved in the retention and elimination of engineered nanoparticles, nanoparticle size is known to play a key role [64, 65]. It is reported that 40–90% (injected dose, ID) nanoparticles, depending on the types of nanoparticles and animal models used in the studies, are usually retained in RES organs 72 hours after injection and can only be cleared or degraded during a long period of time, often several months [66, 67]. Thus, RES organ uptake is the most common pathway for the body to remove engineered nanoparticles with sizes above 10 nm. In the RES organs, nanoparticles with the size of 50–100 nm preferentially accumulate in liver, while 200–500 nm large nanoparticles can be trapped in spleen through the inter-endothelial cell slits of the spleen [68]. Both cells and physical features of the liver contribute to the uptake and clearance of nanoparticles by liver [64]. Specifically, the physical organization of the vascular network in the liver is likely to be a major contributing factor due to their sizes similar to the liver vascular fenestrations [61]. Kupffer cells, which are the resident macrophages of the liver, are considered to be largely responsible for cellular uptake of nanoparticles [62]. These phagocytes act as “garbage collectors” by engulfing cells debris and foreign material, such as bacteria, viruses, and nanoparticles, then removing them [67]. Notably, because the biological relevance and
functions associated to the iron metabolism and usage, iron oxide nanoparticles can be degraded in RES organs such as spleen, then stored and re-used for generating hemoglobin [69]. However, the slow clearance and dosages needed for diagnosis and drug delivery may expose the RES organs and immune system to the biochemically reactive nanoparticles with the risk of toxic effects and unwanted immune-responses. Equally important to be noted, the slow clearance of the nanoparticle imaging probes or drug delivery carriers prevent the agents from being used in the clinical applications that require repeated or longitudinal imaging or administration of the agents for follow-up imaging and treatment.

Since the uptake of nanoparticles by Kupffer cells, or macrophage in general, is size-dependent in favor of larger nanoparticles [70–71], sub-5 nm nanoparticles therefore have relative low accumulation in RES comparing to the larger nanoparticles, shortening the time needed for the RES clearance. Huang et al. measured the change of MRI signals as an indicator of iron oxide nanoparticles accumulated in the liver of mice i.v. injected with saccharide coated iron oxide nanoparticles with the core sizes of 3 and 20 nm (SIO-3 and SIO-20) in the same animals over three weeks after injection (Figure 3). At the same injected dose (2.5 mg Fe/Kg), MRI contrast of the liver recovered to the level of pre-injection after 3 weeks in the mice receiving 3 nm SIO-3, while only 50% signal recovery was observed in the mice receiving 20 nm SIO-20 with the same coating [27]. The clearance of the thick triblock amphiphilic polymer coated same 20 nm iron oxide nanoparticles (SHP-20) was even slower than SIO-20 coated with a thin layer of oligosaccharides.

3.2. Enable Renal Clearance

In contrary to larger nanoparticles with sizes over 6 nm, sub-5 nm nanoparticles can be readily excreted through the kidney by crossing a unique multiple-layer structure of glomeruli [72, 73]. Compared to the RES, the renal secretion is able to rapidly eliminate > 50% ID of nanoparticles within 24 hours after administration with little effect on the cellular level physiology and metabolism. Thus, the renal clearance substantially minimizes the exposure of nanoparticles to the normal organs and tissues, reducing the potential long-term toxicity of nanoparticles accumulated in RES organs. Therefore, it is of great interest to design and prepare the sub-5 nm nanoparticles with proper physical, chemical and biological properties to control and modulate the renal clearance efficiency and profile to allow sufficient blood retention/circulation time for delivery before being excreted through the kidney.

The efficiency of renal clearance of NPs can be affected by nanoparticle size, surface chemistry and charge, shape, and particle density [74, 75]. Various types of size-controllable renal-clearable nanoparticles, including those listed in Table 1, have been developed and tested in murine models and non-human primate models to understand their filtration through glomerular filtration membrane (GFM) for efficient clearance. The general observation is that glomerular filtration is size-dependent due to the ultrastructure of the glomerulus. For instance, the cysteine-coated quantum dots showed a 72% increase (from 43.65 to 75.13% ID) in their renal clearance efficiency (measured 4 h after i.v. administration) with the size decreased by 1.16 nm (from 5.52 to 4.36 nm) [76]. A similar result was observed 48 hours after administration of PEG-coated silica NPs, where the renal
clearance increase (from 64% ID to 73% ID) was caused by 2.7 nm decrease (from 6.0 to 3.3 nm) of their size [51]. Gold nanoparticles also demonstrated size-dependent renal clearance. 24 hours after injection, the renal clearance efficiency increased from 4 to 42% ID and 52.5% ID and 51.6% ID as Au nanoparticles sizes decreased from 6 nm to 3 nm and 2.4 nm and 1 nm, respectively [77]. However, at the sub-nanometer scale, the renal clearance of glutathione-coated gold nanocluster Au_{18}, Au_{15} and Au_{10-11} were 4–9 times slower comparing to Au_{25} (~1.0 nm) [73]. These observations, as illustrated in Figure 4, indicated that nanoparticles with core sizes within the range of 2–6 nm have increased renal clearance efficiency when decreasing sizes, attributed from the interactions between sub-5 nm nanoparticles with the podocytes and glomerular basement membrane [78]. The interaction between nanoparticles and the endothelial glycocalyx is gradually increased when the particle size falls into the sub-nanometer range (≤1 nm), resulting in the decreased renal clearance efficiency due to particles trapped in the kidney.

The kidney filtration and excretion of sub-5 nm imaging enable nanoparticles, such as Au, Ag, Gd and iron oxide, can be monitored in live animals by non-invasive imaging methods at different time points of renal clearance [27, 43, 72–75]. Worth noting, the dynamic imaging and monitoring the renal filtration and clearance of the designed nanoparticle imaging probes with clinical imaging modalities, i.e., CT or MRI, as shown in Figure 4C and D, offers potential applications of sub-5 nm nanoparticles coupled with non-invasive imaging tools to diagnose and assess the kidney functions and diseases.

4. Improved Targeting and Delivery of Nanoparticles to Diseased Tissues

4.1. Promoting the Enhanced Permeability and Retention (EPR) Effect of Nanoparticles

The enhanced permeability and retention (EPR) effect mediated by leaky vasculatures of a tumor or inflamed tissue affected by a disease is widely considered to be a major driving force for nanoparticles to reach and accumulate in the diseased tissue through either passive or active targeting [79]. Various strategies have been explored to take advantage of the EPR effect for increased tumor specific delivery [80, 81], which, in a part, led to the current consensus on the ideal size range of 10 to 200 nm for making nanomaterials for in vivo applications. While increasing the size of nanoparticles promotes the EPR effect by causing longer retention time in the tumor after nanoparticles extravasating from leaky tumor blood vessels, it negatively impacts on the nanoparticles biological properties by introducing other complications (e.g., systematic toxicity, poor clearance and degradation). Importantly, larger nanoparticles are not efficient in permeating through leaky blood vessels, which is an important and the first step in the EPR effect.

Using sub-5 nm uIONP with a 3 nm core size, Wang et al. demonstrated a new and alternative strategy to promote the EPR effect for improving nanoparticle delivery and intratumoral distribution [82]. The authors proposed a mechanism of “bidirectionally” exerting the EPR effect by facilitating more efficient nanoparticle extravasation out of tumor blood vessels to enhance permeability with sub-5 nm nanoparticles, while maintaining or promoting the retention of delivered nanoparticles in the tumor tissue when nanoparticles form clusters due to tight interstitial space or triggered by the tumor microenvironment specific physiological conditions, such as the acidic pH condition (Figure 5A). The greater
tumor accumulation of uIONPs in orthotopic 4T1 mouse model of breast cancer is attributed to more efficient nanoparticle extravasation from tumor vessels and deeper penetration into tumor tissue evidenced by two-photon microscopy (Figure 5B). Given dual T₁-T₂ MRI contrast that is switchable depending on the form of uIONP dispersed in the biological environment, MRI of mice bearing 4T1 breast tumors receiving i.v. injected uIONP showed bright T₁-weighted enhancement in the tumor in the first a few hours after uIONP administration (Figure 5C), suggesting the presence of uIONP in the tumor vasculature. At the later time point (24 hours after injection) T₁ contrast diminished as dark T₂ contrast emerged within the tumor, indicating the uIONPs assemble to large clusters with strong T₂ contrast effect. Easy extravasation of uIONP from the tumor leaky vessels accompanied by the less “wash-out” from the tumor as the result of tissue environment triggered self-assembling/clustering in the tumor interstitial or in the cells therefore enhances the EPR driven passive targeting in tumors and were delivered into the tumor center with a high efficiency compared to that of the particles with large sizes (i.e., 10 and 20 nm).

4.2. Exerting Ligand-Mediated Active Targeting

One of the major advantages of nanomaterials in biomedical applications is to integrate and then deliver various functions and capabilities through the surface chemistry. “Active targeting” based on the ligand-target affinity is a common strategy to precisely deliver nanoparticle imaging probes or drug carriers to the diseased tissue after systemic administration [83, 84]. Various cell specific molecular entities that are over expressed in diseased tissues, such as receptors, are used as biomarkers for targeted imaging and drug delivery with development of high affinity targeting ligands coupled on the selected nanoparticles. Ideally, the biomarker specific active targeting should enable quantitatively imaging biomarkers with ligand functionalized nanoparticle probes for diagnosis, monitoring disease progression and treatment responses, and directing therapeutics to the targeted diseased tissue. However, such ligand-mediated active targeting is inevitably taking place with prerequisite “passive targeting”, driven by the EPR effect.

With most of the studies using nanoparticles in the size range of 10–200 nm, it has been shown that the EPR driven passive targeting plays a dominating role over active targeting, leading to the question and debate on whether the active targeting strategy is sufficiently effective or even necessary [85–87]. Importantly, the presence of EPR-mediated passive targeting and accumulation of off-targeted nanoparticles lead to intrinsic “noise” background that interferes quantitative imaging of biomarkers and delivery of biomarker targeting nanoparticles. Achieving a high level of active targeting to enhance “signal-to-noise ratio” is essential to address this key requirement for targeted therapy by precision medicine.

The recent study by Xu et al. showed that under the same delivery route and tumor microenvironment, using the sub-5 nm nanoparticles, such as 3-nm core size uIONPs, for ligand mediated delivery of imaging capability can significantly improve the efficiency and level of active targeting than that of the larger nanoparticles [88]. The authors compared transferrin (Tf) conjugated uIONP and uIONP without ligand that were co-injected in the same animals bearing the tumors (4T1 breast cancer) with over expression of transferrin receptors (TfR). By labeling two types of nanoparticles with different fluorescent dyes (i.e., FITC and TRITC), co-injected active and passive targeting uIONPs delivered and
accumulated in the same tumors can be readily tracked and quantified by multi-photon imaging (Figure 6A and B). It was found that there was over 6-fold increase in active targeting facilitated tumor accumulation of uIONPs, compared to that of passive targeting ones. Ligand mediated active targeting led to a time dependent accumulation of uIONPs with deeper tumor penetration and prolonged tumor retention time. 24 hours after the injection, more than 75% of uIONPs accumulated in the tumors were found to be active targeting, increasing from ~50% at the early time point (1 hour), while passive targeting uIONP accumulated in the tumors reduced from 50 to 25% as shown in Figure 6C. Thus, the enhanced active targeting by uIONPs can be attributed to the unique sub-5 nm size as majority off-targeted uIONPs were readily cleared from the tumor by intravasation back into tumor blood vessels. The size-dependent tumoral clearance of unbound off-targeted nanoparticles was further supported by their experiments of comparing uIONPs with 30 nm IONPs under the same experimental conditions. The results showed that the active targeting only provided a marginal advantage than passive targeting in tumoral uptake and accumulation of 30 nm IONPs (Figure 6C). The findings support that the ligand mediated active targeting strategy is a valid approach to the targeted delivery of nanoparticles to the tumors under the conditions that can reduce non-specific accumulation of non-targeting or off-targeted nanoparticles, in this case, by reducing the nanoparticle size to sub-5 nm.

5. Improved Properties and Functionalities

5.1. High Payload and Ligand Density

One of the other major interests in nanomedicine is engineering nanoparticles with higher efficiency in delivering therapeutics to diseased tissues in most cases with high specificity and affinity to targeted markers [89]. Nanoparticles with smaller sizes, e.g., sub-5 nm, inherently possess higher surface-to-volume ratios than their counterparts with larger sizes, such as commonly used nanoparticles with 10 to 200 nm diameters, allowing for: (1) carrying more drug molecules in unit weight of nanoparticles to improve the loading efficiency, and (2) conjugating more targeting ligands per unit weight of nanoparticles for effectively targeting biomarkers which enhances the chance of binding between nanomaterials and biomarkers. Based on the relationship of the volume and surface area per unit, the reduced radius of nanoparticles should result in a proportional increase of the surface area for a given weight of sphere nanoparticles with same composition. For example, the total surface area per milligram of IONPs with a 5 nm core size will be twice higher than that of IONPs with the radius twice larger (10 nm). Despite the various mechanisms and approaches for loading payload molecules to nanoparticles, such as hydrophilic/hydrophobic interactions [90, 91], electrostatic attractions [92, 93] and covalent bonding [94, 95], the surface-to-volume ratio of nanoparticles is an important factor determining the loading efficiency of payloads. When comparing the efficiency of loading doxorubicin (DOX) drug molecules, IONPs with different core diameters of 3.5, 10 and 20 nm with same poly(ethylene glycol)-block-allyl glycidyl ether (PEG-b-AGE) coating (Figure 7A to C) [96] exhibited loading efficiencies (DOX/IONP, wt%) of 42.5 ± 2.1%, 21.4 ± 1.8% and 12.8 ± 1.5%, respectively. These PEG-b-AGE polymer coated 3.5, 10 and 20 nm IONPs were mono-dispersed in water with hydrodynamic diameters of 10.3 ± 0.5, 22.4 ± 0.6 and 30.5 ± 1.6 nm (Figure 7D), while demonstrating similar release profiles in an acidic environment.
at pH = 5.0 due to the protonation of the doxorubicin molecule changing the polarity and hydrophobicity of the drug molecules leading to the release of DOX encapsulated in the coating layer (Figure 7E).

Another advantage of smaller nanoparticles with higher surface-to-volume ratios is their capacity for the higher surface ligand density. When surface/volume ratio increases, the number of active sites is increased, suggesting that, by using smaller IONPs (less than 5 nm diameter) to carry ligands, the process could become more efficient [97]. As ligand mediated active targeting to biomarkers by nanoparticles is dependent on the binding between targeting ligands on the surfaces of nanoparticles and targeted tissue or cell surface biomarkers, bearing more ligands on the unit surface area may increase the probability of targeting ligands interacting with the targeted biomarker, thus enhancing targeting efficiency and affinity of nanoparticles to the biomarkers. Using cyclo-(Arg-Gly-Asp-D-Phe-Cys) (RGD), a peptide widely used as the ligand to target the integrin over expressed in many tumors [98, 99], to functionalize PEG-b-AGE polymer coated IONPs of different core sizes, one can measure the number of RGD peptides conjugated to each IONP by measuring the number of amine groups on the surface of IONPs before and after conjugation [100]. It was found that the amount of RGD ligands conjugated to IONPs was 1.32 ± 0.24 mmol/g Fe for uIONPs with a core size of 3 nm comparing to 0.38 ± 0.05 and 0.15 ± 0.01 mmol/g Fe for IONPs with 10 and 20 nm core diameters, respectively (Figure 7F). The high loading with sub-5 nm nanoparticles can be also applied to carry more signal emitting molecules, such as near infrared (NIR) dye Cy5.5 or PET sensitive $^{64}$Cu chelate for PET, for multi-modal imaging to enhance the sensitivity of the detection [32, 51].

5.2. Size-Dependent Novel Imaging Capabilities

Apart from the improvements in chemical, biological and physiological properties with sub-5 nm nanoparticles discussed, the down-size-related unique physical properties were also observed and further explored for imaging applications. Specifically, magnetic iron oxide nanoparticles are known for mainly generating $T_2$-weighted “darkening” contrast in MRI due to drastically shortening the transvers relaxation time ($T_2$) of water molecules in surrounding tissue and strong magnetic susceptibility ($T_2^*$) effect [101, 102]. However, the hypointense or negative contrast in diagnostic MRI has a number of drawbacks, including: poor contrast when used to delineate the abnormalities in the areas that have low background signals such as liver, causing “blooming” artifact and partial volume effect due to the strong susceptibility and diphase effects; co-founding hypointense signal from other materials in the tissue. Furthermore, the strong $T_2$ effect from large sized IONPs may not be translated to better negative contrast in the $T_2$- and $T_2^*$-weighted imaging, since the contrast may not improve further as the level of signal decrease cannot get beyond voiding the signal from its original level [103]. Therefore, there has been significant effort in developing magnetic nanoparticles that can enhance specifically longitudinal relaxation time ($T_1$) of the affected water molecules to generate bright or positive $T_1$-weighted contrast [104, 105]. Since magnetic iron oxide nanoparticles possess both $T_1$ and $T_2$ contrast enhancing effects with the $T_2$ effect being dominant and strongly size-dependent, one rational strategy is to reduce the $T_2$ and susceptibility effect while retaining or improving $T_1$ contrast effect by reducing the particle size to sub-5 nm while making the nanoparticle surface with the thin and high
hydration coating [27]. This strategy is supported by the finding that the relaxivity $r_1$, or $1/T_1$ at a given concentration of specific contrast agent, is strongly dependent on the total surface area of the nanoparticles [106]. The larger surface area of sub-5 nm iron oxide nanoparticles facilitate high probability and close interactions of surrounding water molecules with the nanoparticles. Moreover, the magnetic moment, which contributes strongly to $T_2$ effect, decreases rapidly with the core size decreases [107], leading to increase of $r_1/r_2$ ratio in favor of producing the $T_1$ contrast. Accordingly, Kim et al. synthesized sub-5 nm extremely small-sized iron oxide nanoparticles (ESIONs) with different size via the thermal decomposition process [26]. They demonstrated that the magnetization at room temperature exhibited size-dependent decrease as the core size of the ESIONs decrease from 3.7 nm to 2.2 nm, However, the $r_1$ value was improved from 2.37 mM$^{-1}$s$^{-1}$ to 4.78 mM$^{-1}$s$^{-1}$ (Figure 8A). Because low magnetic moment induces weak magnetic inhomogeneity around the particles, ESIONs with 2.2 nm core size has lower $T_2$ relaxivity and lower $r_2/r_1$ ratio as compared to the larger sized particles, leading to strong $T_1$-weighted contrast for MRI visualization of vasculature with high resolution (0.2 mm). The $r_1$ value of 8.2 mM$^{-1}$ s$^{-1}$ from the Fe$_3$O$_4$ nanoparticles with a 1.7 nm core size reported by Wang et al. is even higher than that of the small lanthanide chelate molecule Gd-DTPA ($r_1$=4.8 mM$^{-1}$ s$^{-1}$) widely used clinical diagnostic imaging [105]. At the $r_1/r_2$ ratio of 0.3, the SIO-3 (or uONP) with 3-nm core size developed by Huang et al. also demonstrated the size-dependent changes in the MRI contrast effect (Figure 8B) and $T_1$-weighted contrast enhancement of vascular systems over prolonged period of time.

It should be noted that carefully designing and modulating the size and coating surface of the sub-5 nm iron oxide nanoparticles can tune the MRI contrast towards more or less $T_1$- or $T_2$-weighted enhancement as described in the recent investigation [106]. Importantly, the $T_1$ or $T_2$ contrast enhancement generated from the rationally designed magnetic nanoparticles has strong dependence on the biological environment or the compartments where the contrast agents accumulate and dispersed. Therefore, it is advantageous to use the sub-5 nm iron oxide nanoparticles as imaging agents to probe the tissue environment or report the delivery, distribution and clearance of the imaging and theranostic nanoparticles over time in vivo by non-invasive MRI methods.

5. Future Directions

While the development of sub-5 nm nanoparticles was initially motivated by the pursuit of safer and more biocompatible nanomaterials for in vivo biomedical applications, it has opened the new opportunities for gaining better understanding of fundamental mechanisms and processes involved in the interactions between engineered nanomaterials and biological systems and accelerating the clinical translation of nanomaterial based imaging and theranostics. Both are major directions for further development of nanomedicine. Increasing studies and new findings on sub-5 nm nanoparticles, as highlighted in this review, clearly demonstrated the advantages and promises of this new class of nanomaterials in addressing several major challenges in current nanomedicine, especially in poor efficiency in delivering imaging and theranostics agents and slow clearance associated safety concerns. Clinical acceptance and commercialization of developed nanomaterials significantly depend on whether a nanomaterial platform can successfully overcome these limitations.
For future development of sub-5 nm nanoparticles, efforts can be made in a number of immediate areas. First of all, renal clearance of the sub-5 nm nanoparticles remains to be fully understood. While a great deal of excellent studies done by Zheng J. et al. and others [30, 43, 70–73] have paved way for the systematical investigation of this complex process, it is also recognized that the differences of materials and surface properties can significantly change the renal clearance profile and time. Further studies on interactions and correlation of renal and vascular functions with well characterized sub-5 nm nanoparticles and their surface properties may allow for elucidating the controlling factors of this process. The information obtained in such studies can be important for ones to tune the appropriate renal clearance time and efficiency to couple with the blood retention time of nanoparticles in order to maximize the delivery. In addition, nanoparticles are known for inducing immunogenicity and immune responses in systemic delivery. The side effect or possible impact of imaging and theranostics nanoparticles on immune systems and immune-therapy are active research topics. Currently, there is little report on whether and how sub-5 nm nanoparticles may play a role. Sub-5 nm nanoparticles have shown to have size advantages in penetrating tumor tissue and less RES uptake. The immune cells, such as macrophages, typically prefer to clear the nanoparticles with larger sizes from the site of accumulation [107]. Therefore, sub-5 nm nanoparticle imaging probes and drug carriers may escape from scavenger macrophages with more efficient delivery to the diseased tissue, such as tumors. In this regard, it is expected that there will be great interests and development in using sub-5 nm nanoparticles as a platform to develop multimodal imaging agents, such as PET/MRI or integrated theranostics such as CT/MRI/photoacoustic imaging with photodynamic or radiation therapy.

**Funding Support**

This study is supported in part by NIH grants (1U01CA19891301, R01CA154846-04 and U01CA151810-05) to HM and LY.

**References**


Figure 1.
TEM images of 2.2 and 3.7 nm iron oxide nanoparticles (A). The size of ESIONs can be controlled by changing aging temperature and solvents. XRD patterns show resulted ESIONs with maghemite ($\gamma$-Fe$_2$O$_3$; JCPDS no. 39–1346) crystal structure (B) and size-dependent magnetization. Images are adopted with permission from the reference 26. Copyright © 2011, American Chemical Society. The size of AgS nanoparticles can be manipulated by changing the reaction time and temperature (C) demonstrated by Hsu et al. [44]. Images are adopted with permission from reference 44.
Figure 2.
Examples of surface coating approaches to render sub-5 nm nanoparticle water soluble and stable. A: Anchoring zwitterionic glutathione molecules to the nanoparticle surface via chemical bonds [29, 30, 33, 44]. B: Forming a thin layer of oligosaccharide on the nanoparticle surface via “in situ polymerization” of glucose. C: Photographs of oligosaccharide solutions under normal light (upper) and UV light (bottom) at the different stages during the polymerization reaction (adopted from adapted with permission from the reference 27).
Figure 3.
Clearance studies of i.v. administered 3 nm SIO nanoparticles at 2.5 mg Fe/kg in BALB/c mice. (A) Pseudo colored T₂-weighted MR images following up the clearance of nanoparticles in liver (L) and spleen (S). The corresponding signal change for (B) liver and (C) spleen in T₂-maps (n=3). Adapted from the reference 27 with permission.
Figure 4.
Sub-5 nm nanoparticles are renally clearable with clearance efficiency controllable based on the size and surface properties. A: The renal clearance efficiencies of Au$_{10-11}$, Au$_{15}$, Au$_{18}$ and Au$_{25}$, 1.7 nm (Au$_{201}$), 2.5 nm (Au$_{640}$) and 6 nm (Au$_{8856}$) GS-AuNPs 24h post injection over the number of gold atoms. Below Au$_{25}$, the renal clearance efficiency exponentially decreased with the decrease of the number of gold atoms in the NPs (adopted with permission from reference 74). B: Illustration of the components of glomerulus. The glomerular filtration membrane is composed of multiple layers: endothelial glycocalyx, endothelial cell, glomerular basement membrane (GBM) and podocyte (adapted from with permission the reference 74). C: T$_1$-weighted (top row) and T$_2$-weighted MRI shows the time-dependent contrast changes in the bladder of a mouse injected with 3-nm core uIONP, as renally excreted uIONP slowly filling the bladder (adopted from reference 27 with permission). D: 3D volume-rendered CT images of a mouse injected with GSH coated
Ag$_2$S nanoparticles with an average core diameter of 3.1 nm. Green circles indicate the bladder which was filled by CT-sensitive Ag$_2$S nanoparticles 5 minutes after the injection. Hearts and kidneys are labeled H and K, respectively (adopted from the reference 43 with permission from the authors).
Figure 5.
The possible mechanism of improving EPR driven delivery of nanoparticles at sub-5 nm scale (A). Multiphoton imaging revealed size-dependent tumor tissue penetration and intratumoral distribution in favor to sub-5 nm uIONP (green) comparing to larger 20-nm core IONP (red) co-injected to the mice bearing 4T1 breast tumors (B). MRI with compartment specific $T_1$-$T_2$ contrast switching from uIONP and iron quantification from the tissue section confirmed size-dependent increase of tumor uptake (C). Adopted with permission from reference 82, Copyright © 2018, American Chemical Society.
Figure 6.

3D and cross-sectional re-construction of multiphoton microscopic images taken from an 8 mm³ tumor tissue block collected from 4T1 tumor-bearing mice co-injected with active targeting FITC-Tf-ulIONPs (green) and non-targeting TRITC-ulIONPs (red) at different time points (with the selected cross-sections) (A). Tumor collagen was visualized using second harmonic generation (SHG), and presented as bright signals in a grayscale setting. The scale bar for all images is 50 μm. Confocal microscopic images of tissue sections containing FITC-Tf-ulIONPs (green) and non-targeting TRITC-ulIONPs (red) and DAPI was used to stain the nuclei, and H&E stained for tissue morphology (B). Time-dependent changes of actively targeted ulIONPs based on multiphoton imaging quantification showed high accumulation of ligand mediated actively targeted ulIONPs comparing to passively targeted controls (C). Adopted from the reference 89 with permission from the authors.
Figure 7.
TEM images of PEG-b-AGE coated IONPs with core diameters of 3.5 (A), 10 (B) and 20 (C) nm and the hydrodynamic diameters of these IONPs in water (D) reveal their stability in aqueous solution. The varied loading efficiencies of DOX onto PEG-b-AGE coated IONPs with 3.5, 10 and 20 nm core diameters decreased at pH 5.0 (E), suggesting IONPs with smaller size have higher payload loading efficiency while having little effect on the payload release. The surface densities of RGD ligands conjugated to each IONP significantly decreased as the core diameters of IONPs increased from 3.5 to 20 nm (F), indicating the higher ligand density on smaller IONPs than that on the larger ones.
Figure 8.
Tuning MRI contrast with sub-5 nm magnetic iron oxide nanoparticles. Magnetization of the nanoparticles is size-dependent with reduced magnetic moment and susceptibility in sub-5 nm scale (A, adopted from the reference 26 with permission, Copyright © 2011, American Chemical Society). Reducing the nanoparticle size leads to high $r_1/r_2$ ratio (B) and enhanced vasculature and kidney by 3-nm SIO-3 in $T_1$-weighted MRI at the field strength of 3T (C), adopted from the reference 27 with permission. By modulating the surface coating, $T_1$ and $T_2$ dual contrast can be optimized as shown in MRI of subcutaneous tumors (D, adopted from the reference 108 with permission, Copyright © 2017, American Chemical Society).
Table 1.
Examples of different types of sub-5 nm nanoparticles developed for *in vivo* applications

<table>
<thead>
<tr>
<th>Name and Materials*</th>
<th>TEM Measured Size (nm)</th>
<th>DLS Measured Size (nm)</th>
<th>Surface Coating</th>
<th>Renal Clearance</th>
<th>Applications</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>GS-AgNP</td>
<td>2.6 ± 0.2</td>
<td>3.1 ± 0.3</td>
<td>Glutathione (GSH)</td>
<td>51.36% ID within 48 h</td>
<td>CT</td>
<td>[29]</td>
</tr>
<tr>
<td>GS-Au/Ag</td>
<td>2.6 ± 0.2</td>
<td>3.1 ± 0.3</td>
<td>GSH</td>
<td>52.99% ID within 48 h</td>
<td>CT/NIR</td>
<td>[29]</td>
</tr>
<tr>
<td>GS-AuNPs</td>
<td>2.5 nm</td>
<td>3.3</td>
<td>GSH</td>
<td>69.0% ID within 2 h</td>
<td>CT/NIR</td>
<td>[30]</td>
</tr>
<tr>
<td>Gly-Cys-Au</td>
<td>2.31</td>
<td>3.12</td>
<td>Gly and Cys</td>
<td>41.6% ID within 24 h</td>
<td>CT/NIR</td>
<td>[31]</td>
</tr>
<tr>
<td>³Cu-NOTA-Au-GSH</td>
<td>2.0 ± 0.4</td>
<td>2.6 ± 0.1</td>
<td>GSH, p-SCN-Bn-NOTA, ³Cu</td>
<td>&gt;75% ID within 24 h</td>
<td>PET/CT/NIR</td>
<td>[32]</td>
</tr>
<tr>
<td>GS-Au</td>
<td>2.6 ± 0.4</td>
<td>3.4 ± 0.4</td>
<td>GSH</td>
<td>52.5 % ID within 24 h</td>
<td>CT/NIR</td>
<td>[33]</td>
</tr>
<tr>
<td>uIONPs</td>
<td>3.2 ± 0.4</td>
<td>4.5 ± 0.6</td>
<td>oligosaccharides</td>
<td>50% ID within 48 h</td>
<td>MRI</td>
<td>[27]</td>
</tr>
<tr>
<td>ZES-SPIONs</td>
<td>3</td>
<td>4.7</td>
<td>glucose</td>
<td>65% ID within 48 h</td>
<td>MRI</td>
<td>[43]</td>
</tr>
<tr>
<td>AgS</td>
<td>3.1 ± 0.5</td>
<td>4.9 ± 1.0</td>
<td>GSH</td>
<td>85 ± 2% within 24 h</td>
<td>CT</td>
<td>[44]</td>
</tr>
<tr>
<td>Au@DTDTDPA-%sup3;In</td>
<td>2.4±0.5</td>
<td>6.6 ± 1.8</td>
<td>%sup3;In-DTDTDPA</td>
<td>64% ID within 24 h</td>
<td>NIR/SPECT/CT</td>
<td>[45]</td>
</tr>
<tr>
<td>AuNPs</td>
<td>2.6</td>
<td>3.0</td>
<td>GSH</td>
<td>&gt;50% ID within 48 h</td>
<td>NIR/SPECT/CT</td>
<td>[46]</td>
</tr>
<tr>
<td>GC-AuNPs</td>
<td>2.3±0.4</td>
<td>2.9 ± 0.3</td>
<td>GSH and cysteamine</td>
<td>40–50% ID within 2 h</td>
<td>CT/NIR</td>
<td>[47]</td>
</tr>
<tr>
<td>Peptide-templated Au</td>
<td>1. ±0.4</td>
<td>~11</td>
<td>GSH and peptide</td>
<td>73 ± 7% within 1 h</td>
<td>CT/NIR</td>
<td>[48]</td>
</tr>
<tr>
<td>QD-Cys</td>
<td>3.02</td>
<td>4.91 ± 0.05</td>
<td>cysteine</td>
<td>50% ID within 4 h</td>
<td>Fluorescence</td>
<td>[49]</td>
</tr>
<tr>
<td>Silica Nanoparticle</td>
<td>NR</td>
<td>6.8</td>
<td>125I-cRADDY-PEG</td>
<td>72% ID in 72 h</td>
<td>PET</td>
<td>[50]</td>
</tr>
<tr>
<td>Cy5 C-Dot (silica)</td>
<td>NR</td>
<td>3.3 ± 0.06</td>
<td>Cy5</td>
<td>73% ID within 48 h</td>
<td>Fluorescence</td>
<td>[51]</td>
</tr>
<tr>
<td>PEG-Cy5 C Dot (silica)</td>
<td>NR</td>
<td>6.0 ± 0.1</td>
<td>PEG-Cy5</td>
<td>64% ID within 48 h</td>
<td>Fluorescence</td>
<td>[52]</td>
</tr>
<tr>
<td>GdOF/Gd2O3</td>
<td>2.1 ± 0.2</td>
<td>4.6 ± 0.2</td>
<td>Poly-acrylic acid</td>
<td>~50% ID within 48 h</td>
<td>MRI</td>
<td>[53]</td>
</tr>
</tbody>
</table>

*Note: ID = Indistinguishable; CT = Computed Tomography; NIR = Near-Infrared; MRI = Magnetic Resonance Imaging; PET = Positron Emission Tomography; SPECT = Single Photon Emission Computerized Tomography; NR = Not Reported.*