ViNAS-Pro Tutorial

Contents

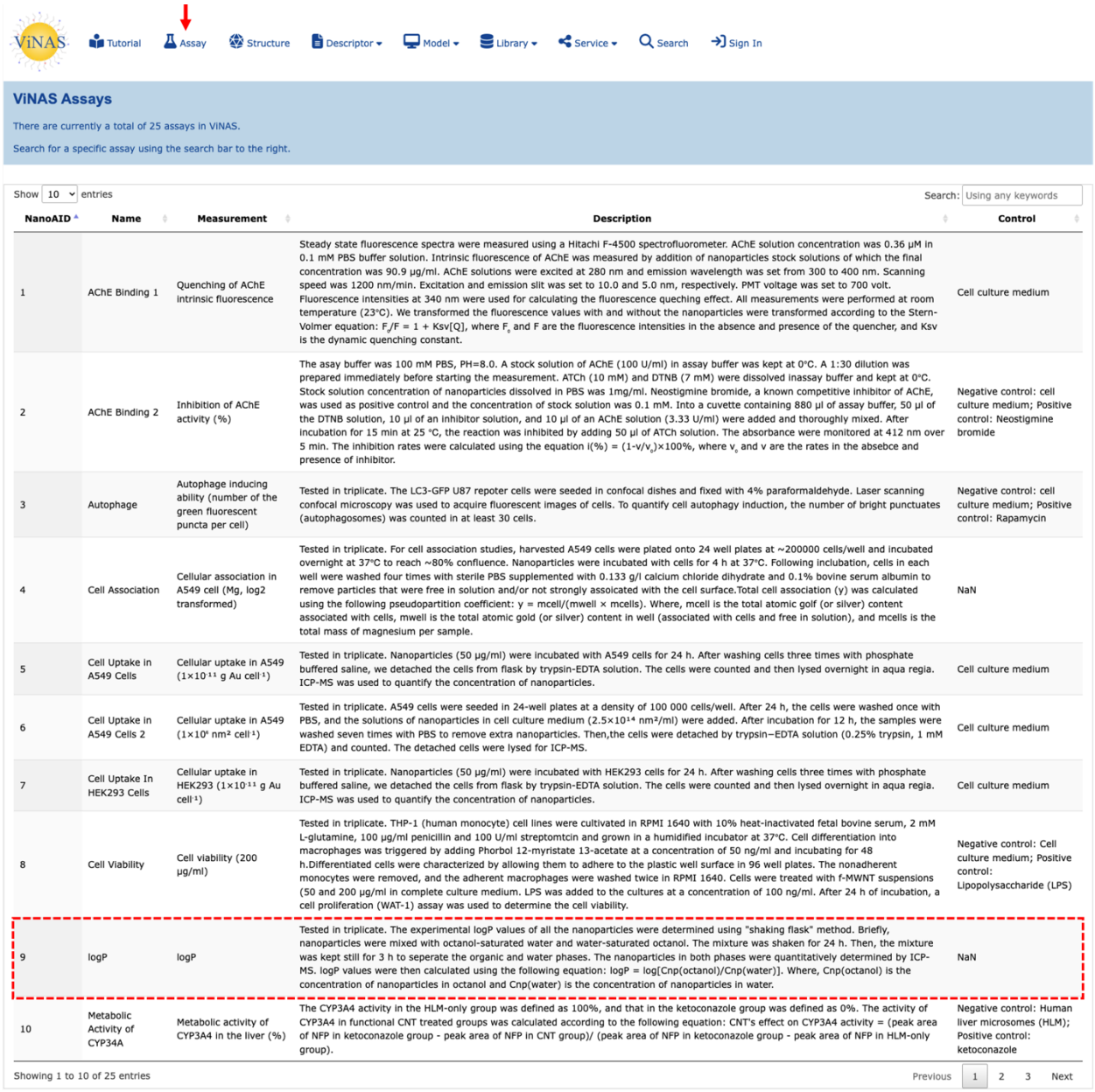
1. Introduction

Virtual Nanomaterial Simulation Professional (ViNAS-Pro, https://vinas-toolbox.com/) is a data-driven nanoinformatics platform. It provides high-quality data, user-friendly modeling tools and endpoint predictions, supporting rational design of new nanomaterials (NMs).

ViNAS-Pro maintains two machine readable databases: the Structure database and the Assay database. The Structure database provides structural information for 13 types of NMs, while the Assay database offers data on the experimentally assessed properties and biological activities of these NMs across 25 different assays. The Descriptor toolkit provides users with modules for data visualization and preprocessing, ensuring the structure diversity of the training data in the machine learning (ML) modeling procedure. The Model toolkit includes two modules: NanoPredictor and AutoNanoML. The NanoPredictor module maintains pre-developed ML models, enabling users to predict specific endpoints for new NMs. The AutoNanoML module provides an interface that allows users to develop their own ML models for various prediction purposes. ViNAS-Pro virtual library provides data analysis, structure data, and endpoint predictions for virtual NMs. Moreover, ViNAS-Pro provides services for data deposit, nanostructure construction, and nanodescriptor calculation.

2. Assay Database

The navigation page of the assay database provides an interactive table that lists the available assays on ViNAS-Pro (**Figure 2.1**). Users can search for a specific assay using keywords and access individual assay records by clicking on entries in the interactive table of the navigation page. The assay record page provides detailed information about a specific assay on ViNAS-Pro. For example, a record for assay nine (NanoAID-9) is shown in **Figure 2.2**. This assay record page includes a figure displaying the activity distributions of NMs tested against NanoAID-9, and an interactive table containing the results of NMs associated with NanoAID-9. Users can download both the assay results and the associated NMs' nanodescriptor data as XLSX files from the assay record page. The endpoint definition, experimental protocol, and related literature are displayed on the assay record page. Furthermore, users can access a specific NM record page (**Figure 2.3**) by clicking on a particular NM within the interactive table on the assay record page.



NanoAID	Name	Measurement	Description	Control
1	AChE Binding 1	Quenching of AChE intrinsic fluorescence	Steady state fluorescence spectra were measured using a Hitachi F-4500 spectrofluorometer. AChE solution concentration was 0.36 μM in 0.1 mM PBS buffer solution. Intrinsic fluorescence of AChE was measured by addition of nanoparticles stock solutions of which the final concentration was 90.9 μg/ml. AChE solutions were excited at 280 nm and emission wavelength was set from 300 to 400 nm. Scanning speed was 1200 nm/min. Excitation and emission slit was set to 10.0 and 5.0 nm, respectively. PMT voltage was set to 700 volt. Fluorescence intensities at 340 nm were used for calculating the fluorescence quenching effect. All measurements were performed at room temperature (23°C). We transformed the fluorescence values with and without the nanoparticles were transformed according to the Stern-Volmer equation: $F_0/F = 1 + Ksv[Q]$, where F_0 and F are the fluorescence intensities in the absence and presence of the quencher, and Ksv is the dynamic quenching constant.	Cell culture medium
2	AChE Binding 2	Inhibition of AChE activity (%)	The asay buffer was 100 mM PBS, PH=8.0. A stock solution of AChE (100 U/ml) in assay buffer was kept at 0°C. A 1:30 dilution was prepared immediately before starting the measurement. ATCh (10 mM) and DTNB (7 mM) were dissolved inassay buffer and kept at 0°C. Stock solution concentration of nanoparticles dissolved in PBS was 1mg/ml. Neostigmine bromide, a known competitive inhibitor of AChE, was used as positive control and the concentration of stock solution was 0.1 mM. Into a cuvette containing 880 μl of assay buffer, 50 μl of the DTNB solution, 10 μl of an inhibitor solution, and 10 μl of an AChE solution (3.33 U/ml) were added and thoroughly mixed. After incubation for 15 min at 25 °C, the reaction was inhibited by adding 50 μl of ATCh solution. The absorbance were monitored at 412 nm over 5 min. The inhibition rates were calculated using the equation i(%) = (1-v/v_0)×100%, where v_0 and v are the rates in the absebce and presence of inhibitor.	Negative control: cell culture medium; Positive control: Neostigmine bromide
3	Autophage	Autophage inducing ability (number of the green fluorescent puncta per cell)	Tested in triplicate. The LC3-GFP U87 repoter cells were seeded in confocal dishes and fixed with 4% paraformaldehyde. Laser scanning confocal microscopy was used to acquire fluorescent images of cells. To quantify cell autophagy induction, the number of bright punctuates (autophagosomes) was counted in at least 30 cells.	Negative control: cell culture medium; Positive control: Rapamycin
4	Cell Association	Cellular association in A549 cell (Mg, log2 transformed)	Tested in triplicate. For cell association studies, harvested A549 cells were plated onto 24 well plates at ~200000 cells/well and incubated overnight at 37°C to reach ~80% confluence. Nanoparticles were incubated with cells for 4 h at 37°C. Following inclubation, cells in each well were washed four times with sterile PBS supplemented with 0.133 g/l calcium chloride dihydrate and 0.1% bovine serum albumin to remove particles that were free in solution and/or not strongly assoicated with the cell surface.Total cell association (y) was calculated using the following pseudopartition coefficient: y = mcell/(mwell × mcells). Where, mcell is the total atomic golf (or silver) content associated with cells, mwell is the total atomic gold (or silver) content in well (associated with cells and free in solution), and mcells is the total mass of magnesium per sample.	NaN
5	Cell Uptake in A549 Cells	Cellular uptake in A549 ($1×10^{-11}$ g Au $cell^{-1}$)	Tested in triplicate. Nanoparticles (50 μg/ml) were incubated with A549 cells for 24 h. After washing cells three times with phosphate buffered saline, we detached the cells from flask by trypsin-EDTA solution. The cells were counted and then lysed overnight in aqua regia. ICP-MS was used to quantify the concentration of nanoparticles.	Cell culture medium
6	Cell Uptake in A549 Cells 2	Cellular uptake in A549 ($1×10^{6}$ nm^2 $cell^{-1}$)	Tested in triplicate. A549 cells were seeded in 24-well plates at a density of 100 000 cells/well. After 24 h, the cells were washed once with PBS, and the solutions of nanoparticles in cell culture medium ($2.5×10^{14}$ nm^2/ml) were added. After incubation for 12 h, the samples were washed seven times with PBS to remove extra nanoparticles. Then,the cells were detached by trypsin−EDTA solution (0.25% trypsin, 1 mM EDTA) and counted. The detached cells were lysed for ICP-MS.	Cell culture medium
7	Cell Uptake In HEK293 Cells	Cellular uptake in HEK293 ($1×10^{-11}$ g Au $cell^{-1}$)	Tested in triplicate. Nanoparticles (50 μg/ml) were incubated with HEK293 cells for 24 h. After washing cells three times with phosphate buffered saline, we detached the cells from flask by trypsin-EDTA solution. The cells were counted and then lysed overnight in aqua regia. ICP-MS was used to quantify the concentration of nanoparticles.	Cell culture medium
8	Cell Viability	Cell viability (200 μg/ml)	Tested in triplicate. THP-1 (human monocyte) cell lines were cultivated in RPMI 1640 with 10% heat-inactivated fetal bovine serum, 2 mM L-glutamine, 100 μg/ml penicillin and 100 U/ml streptomtcin and grown in a humidified incubator at 37°C. Cell differentiation into macrophages was triggered by adding Phorbol 12-myristate 13-acetate at a concentration of 50 ng/ml and incubating for 48 h.Differentiated cells were characterized by allowing them to adhere to the plastic well surface in 96 well plates. The nonadherent monocytes were removed, and the adherent macrophages were washed twice in RPMI 1640. Cells were treated with f-MWNT suspensions (50 and 200 μg/ml in complete culture medium. LPS was added to the cultures at a concentration of 100 ng/ml. After 24 h of incubation, a cell proliferation (WAT-1) assay was used to determine the cell viability.	Negative control: Cell culture medium; Positive control: Lipopolysaccharide (LPS)
9	logP	logP	Tested in triplicate. The experimental logP values of all the nanoparticles were determined using "shaking flask" method. Briefly, nanoparticles were mixed with octanol-saturated water and water-saturated octanol. The mixture was shaken for 24 h. Then, the mixture was kept still for 3 h to seperate the organic and water phases. The nanoparticles in both phases were quantitatively determined by ICP-MS. logP values were then calculated using the following equation: logP = log[Cnp(octanol)/Cnp(water)]. Where, Cnp(octanol) is the concentration of nanoparticles in octanol and Cnp(water) is the concentration of nanoparticles in water.	NaN
10	Metabolic Activity of CYP34A	Metabolic activity of CYP3A4 in the liver (%)	The CYP3A4 activity in the HLM-only group was defined as 100%, and that in the ketoconazole group was defined as 0%. The activity of CYP3A4 in functional CNT treated groups was calculated according to the following equation: CNT's effect on CYP3A4 activity = (peak area of NFP in ketoconazole group - peak area of NFP in CNT group)/ (peak area of NFP in ketoconazole group - peak area of NFP in HLM-only group).	Negative control: Human liver microsomes (HLM); Positive control: ketoconazole

Figure 2.1 The navigation page of the Assay database

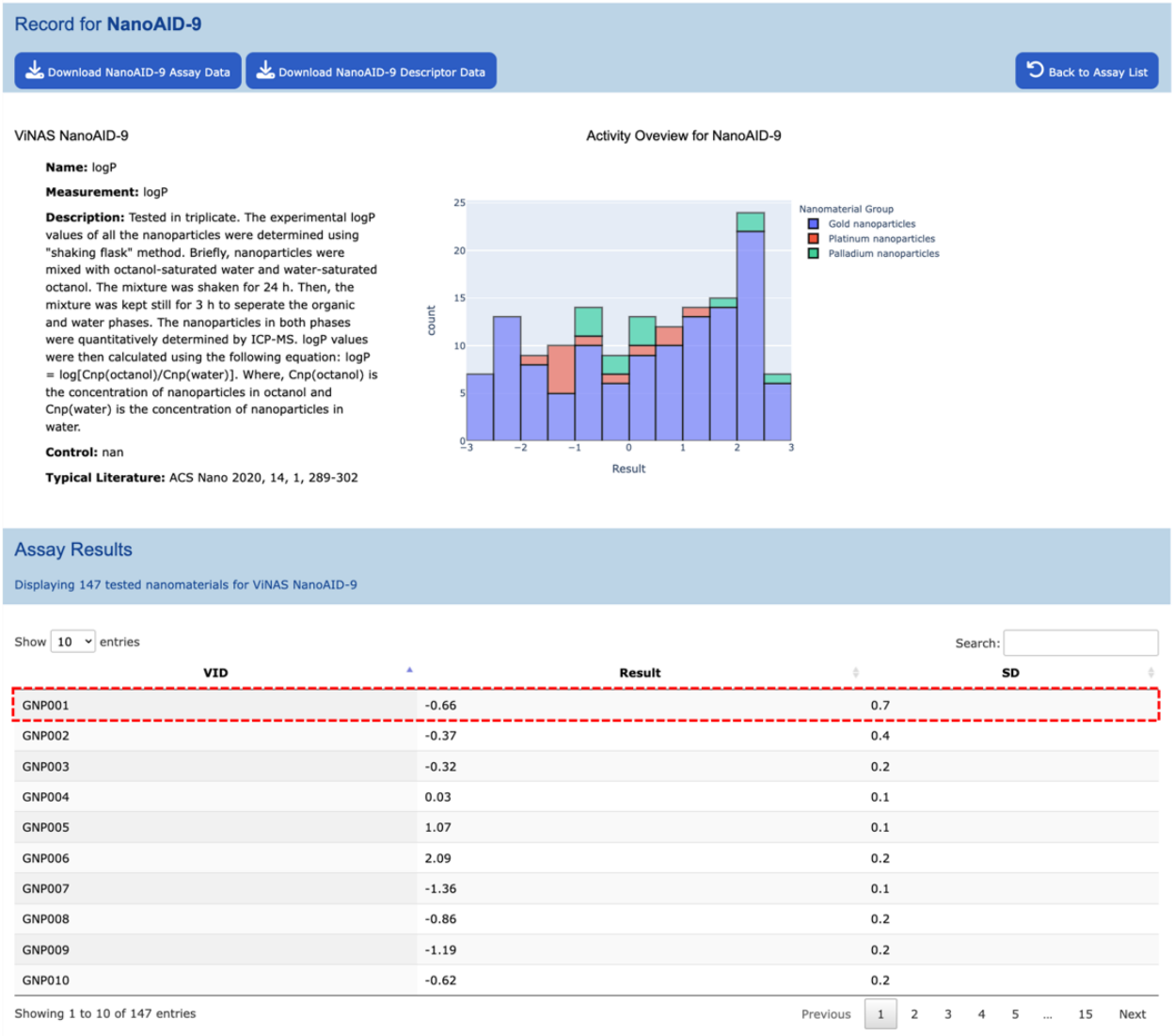


Figure 2.2 The record page for assay 9 (NanoAID-9)

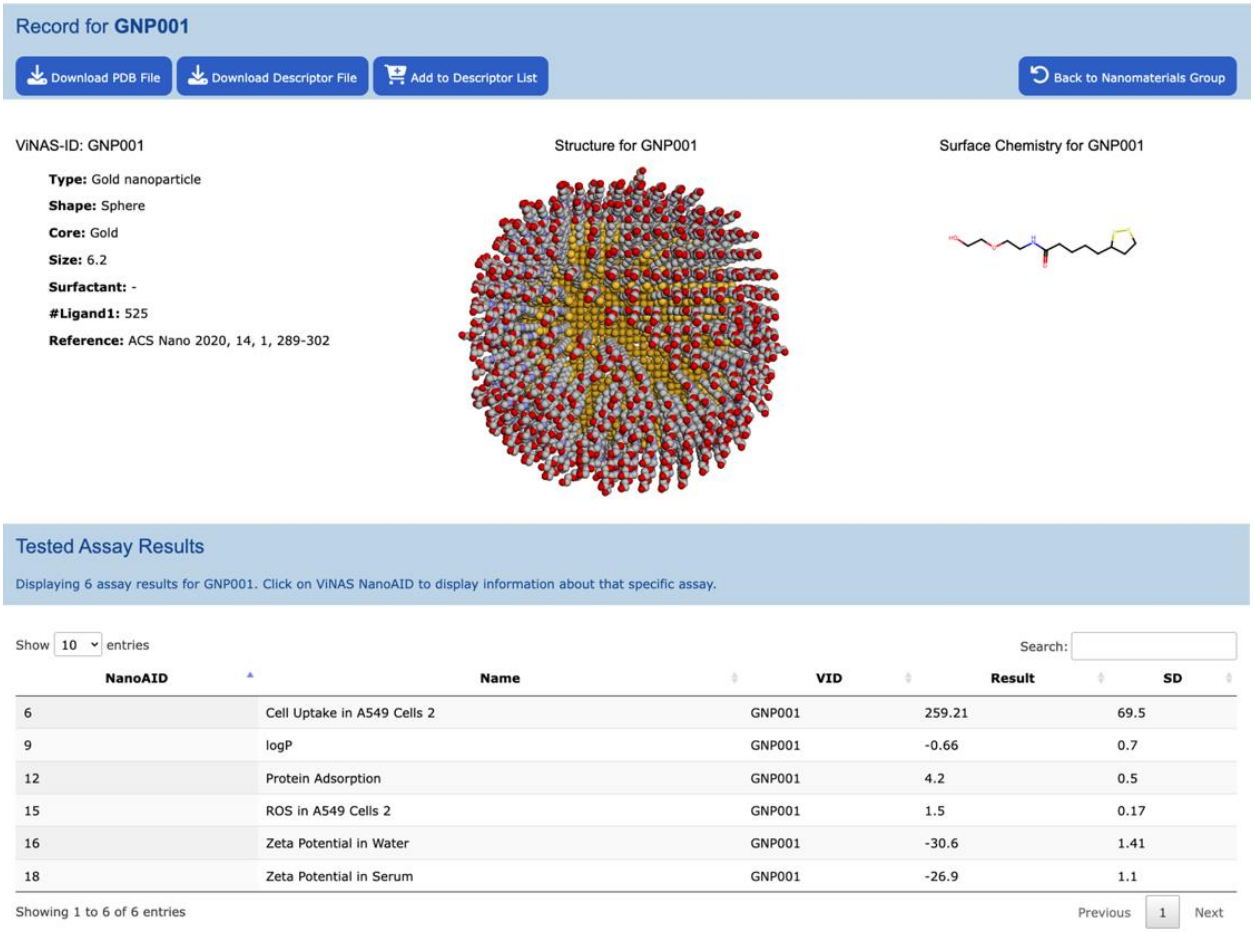


Figure 2.3 The record page for a specific NM

3. Structure Database

The primary navigation page of the structure database provides interactive nanostructures that lists the available NM types on ViNAS-Pro (**Figure 3.1**). Users can access a secondary navigation page of a specific NM type (Figure 2.3) by clicking on a particular NM within the interactive structure on the primary navigation page. Its secondary navigation page provides an interactive table that lists the available NMs (**Figure 3.2**). Users can batch download the structure data as PDB file and nanodescriptor data as XLSX file for associated NM type on the secondary navigation page. Additionally, they can also search for a specific NM record using ViNAS-ID (VID) and access individual NM records by clicking on entries in the interactive table of the secondary navigation page. The NM record page provides detailed information about a specific NM on ViNAS-Pro. For example, **Figure 3.3** shows a record for a gold nanoparticle (GNP) named GNP001 on ViNAS-Pro. GNP001's record page provides its nanostructure figure

rendering in Van der Waals (VDW) format, along with basic structure information such as shape, size, core and ligand. Moreover, it includes an interactive table containing all the assay testing results associated with GNP001. The annotated nanostructure as a PDB file and nanodescriptor data as a XLSX file for GNP001 can also be downloaded from its record page. Furthermore, users can access a specific assay record page (**Figure 3.4**) by clicking on a particular assay within the interactive table on the NM record page.

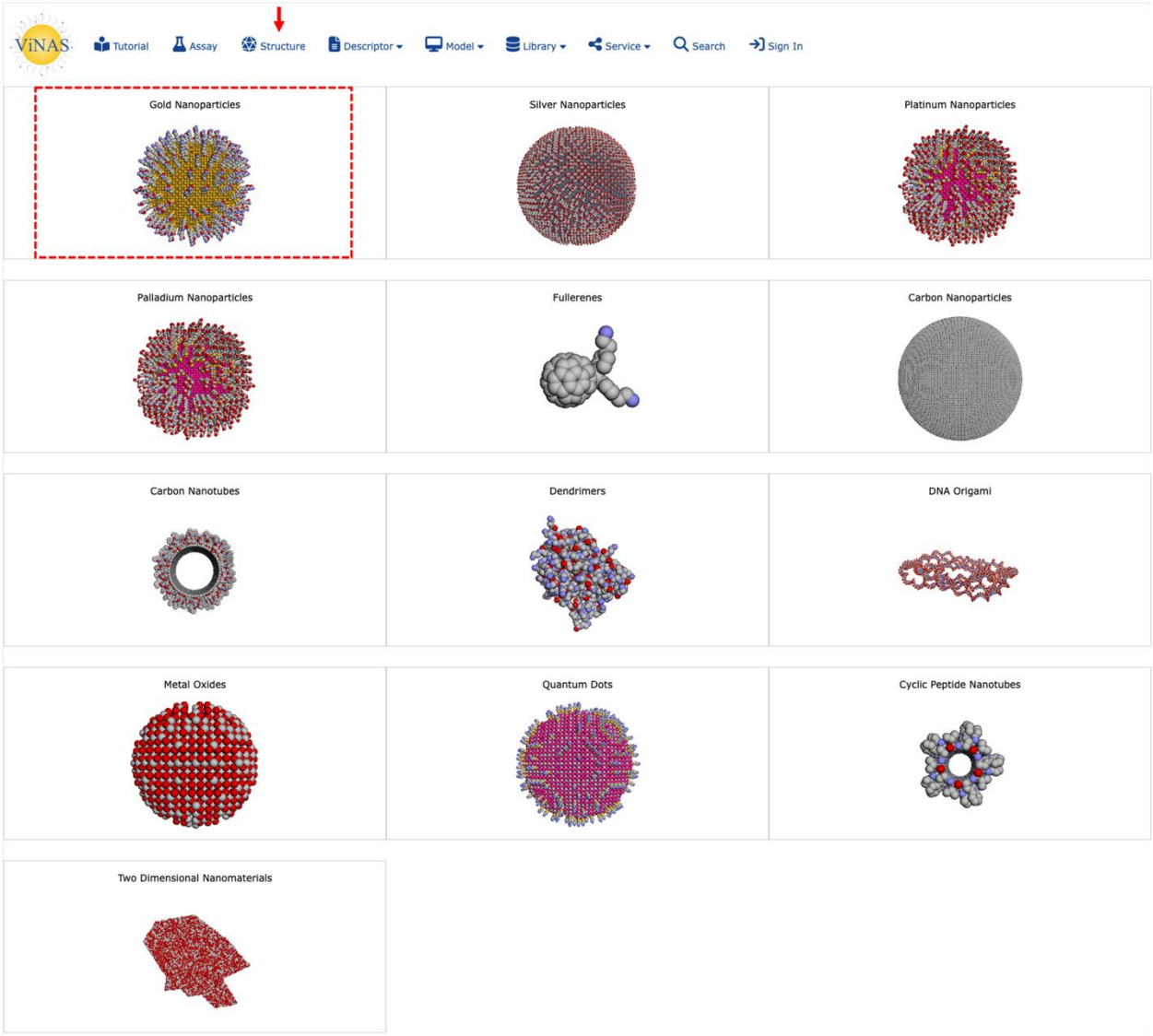


Figure 3.1 The primary navigation page of the Structure database

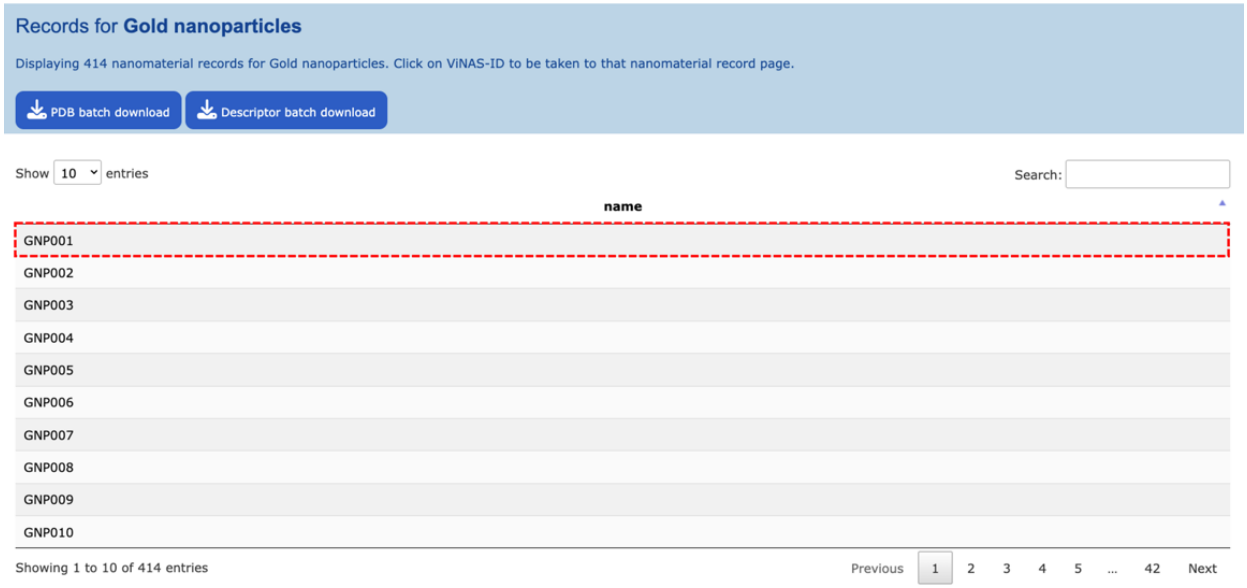


Figure 3.2 The secondary navigation page of the Structure database

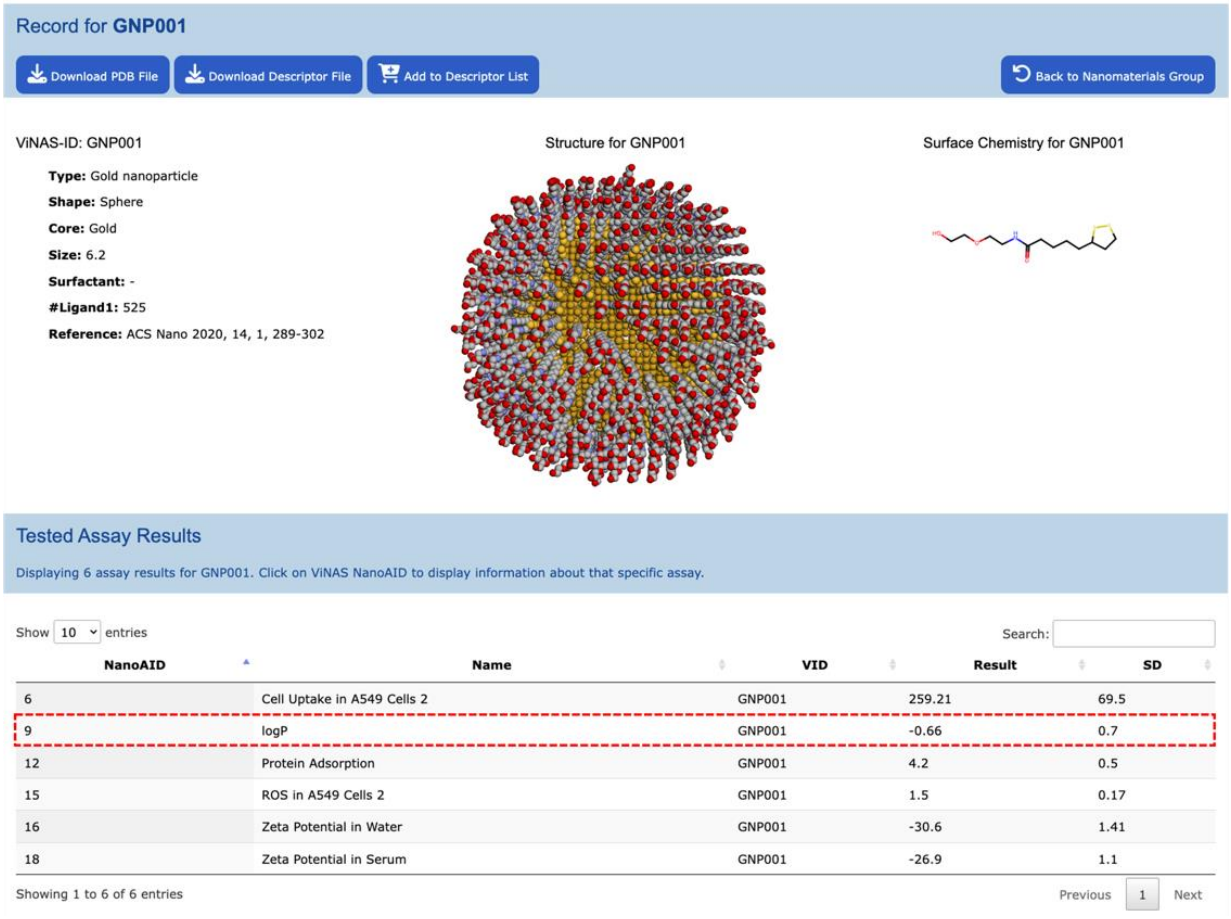


Figure 3.3 The record page for a specific NM (GNP001)

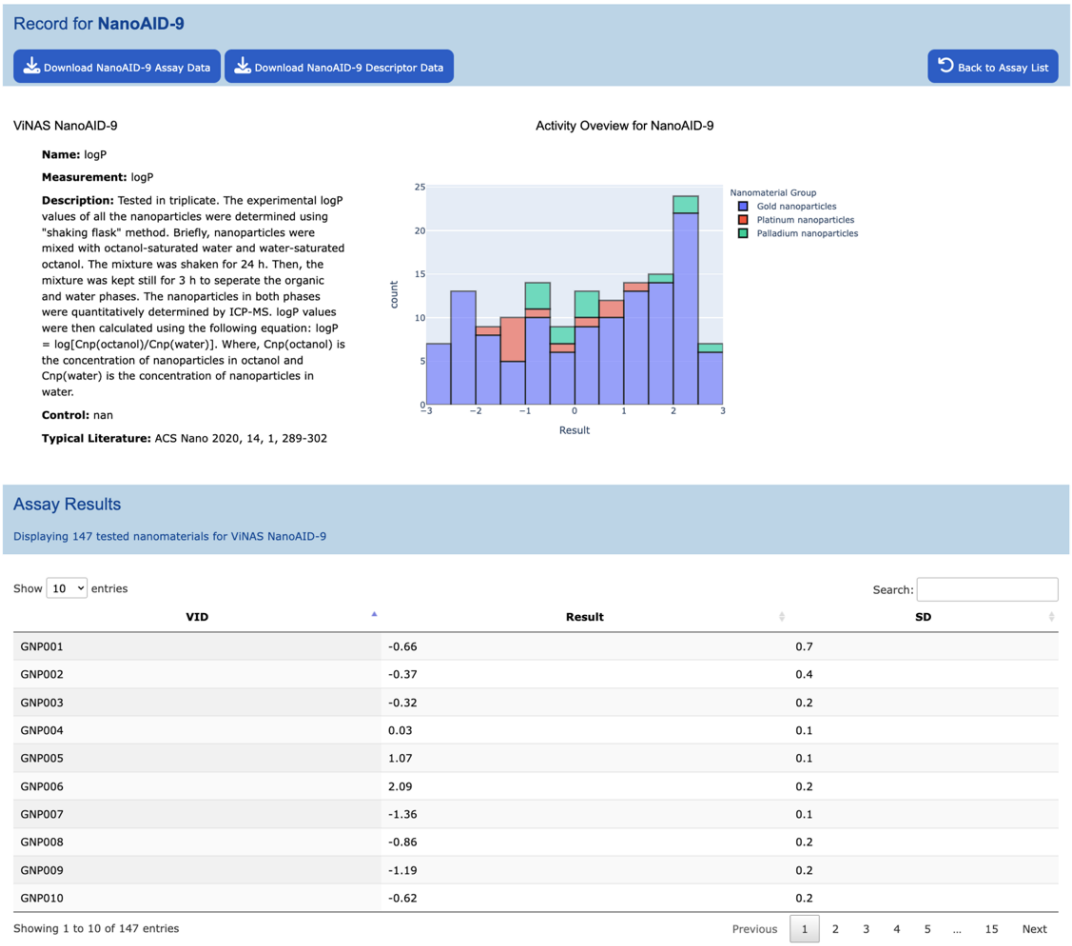


Figure 3.4 The record page for assay 9 (NanoAID-9)

4. Descriptor Toolkit

The Descriptor toolkit allows users to standardize nanodescriptor values using descriptor preprocessing method and analyze the associated NM space using principal component analysis (PCA). The Descriptor List module allows users to analyze the nanodescriptors of target NMs on ViNAS-Pro. Users can selectively add the nanodescriptors of interest of NMs to the Descriptor List interface from the nanostructure record page. For example, the record page for GNP001, provides an interactive function to add its nanodescriptors to the Descriptor List page, as shown in **Figure 4.1**. Subsequently, users can generate a customized descriptor list for specific NMs and submit it for further analysis, following the application of preprocessing functions such as StandardScaler or MinMaxScaler (**Figure 4.2**). The descriptor analysis results from Descriptor List approach are shown on the descriptor analysis page (**Figure 4.3**). Both two-dimensional (2D) and three-dimensional (3D) spaces of NMs are shown by applying PCA to reduce the

dimensionality of nanodescriptors. Each dot represents a NM and provides the NM's coordinates in the corresponding space. The standardized nanodescriptor dataset, along with the 2D and 3D NM space charts, are downloadable on the descriptor analysis page. The Descriptor Upload module allows users to upload their nanodescriptor data for analysis. For example, users can prepare their own nanodescriptor dataset for NMs in XLSX format. They can then submit the nanodescriptor set for analysis (**Figure 4.4**). Similar with Descriptor List module, the descriptor analysis results from Descriptor Upload approach are shown on the descriptor analysis page (**Figure 4.5**).

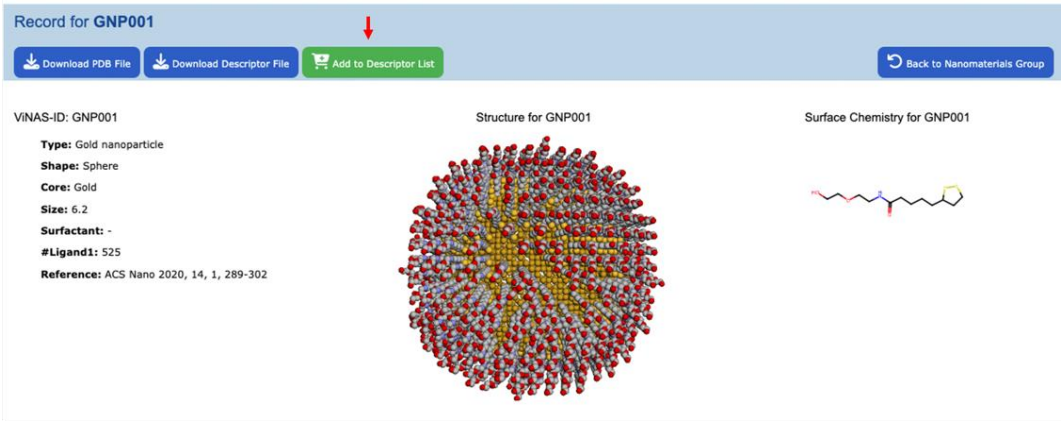
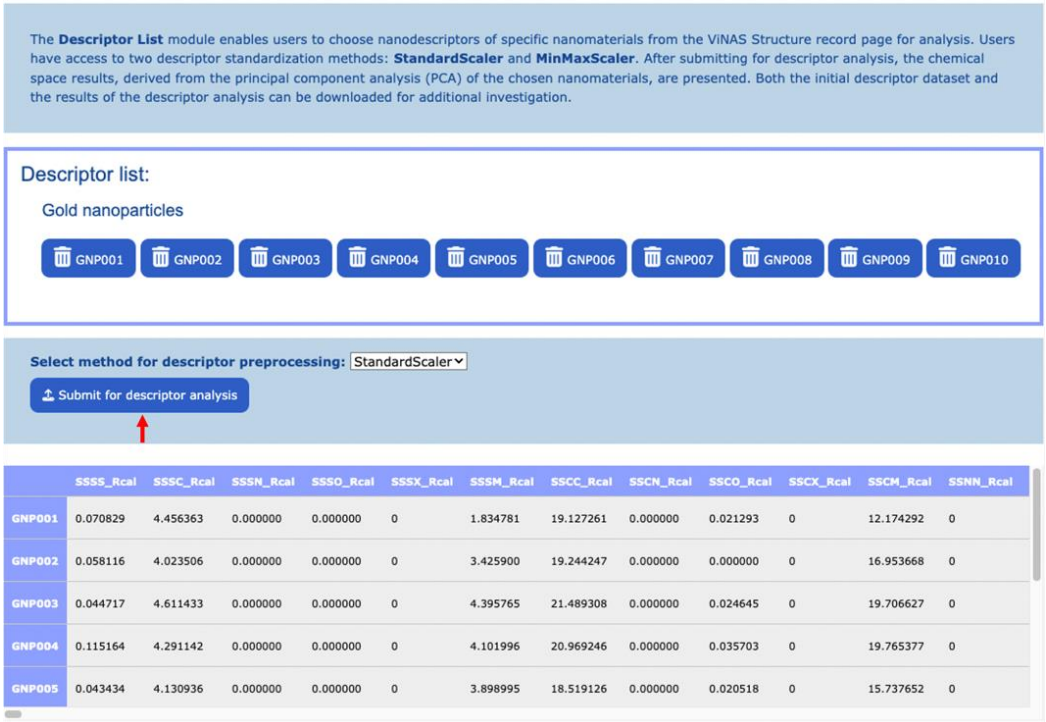


Figure 4.1 Adding nanodescriptors to the Descriptor List module on the GNP001 record page



	SSSS_Rcal	SSSC_Rcal	SSSN_Rcal	SSSO_Rcal	SSSX_Rcal	SSSM_Rcal	SSCC_Rcal	SSCN_Rcal	SSCO_Rcal	SSCX_Rcal	SSCM_Rcal	SSNN_Rcal
GNP001	0.070829	4.456363	0.000000	0.000000	0	1.834781	19.127261	0.000000	0.021293	0	12.174292	0
GNP002	0.058116	4.023506	0.000000	0.000000	0	3.425900	19.244247	0.000000	0.000000	0	16.953668	0
GNP003	0.044717	4.611433	0.000000	0.000000	0	4.395765	21.489308	0.000000	0.024645	0	19.706627	0
GNP004	0.115164	4.291142	0.000000	0.000000	0	4.101996	20.969246	0.000000	0.035703	0	19.765377	0
GNP005	0.043434	4.130936	0.000000	0.000000	0	3.898995	18.519126	0.000000	0.020518	0	15.737652	0

Figure 4.2 Nanodescriptors analysis through the Descriptor List module

Figure 4.3 Nanodescriptors analysis results from the Descriptor List approach

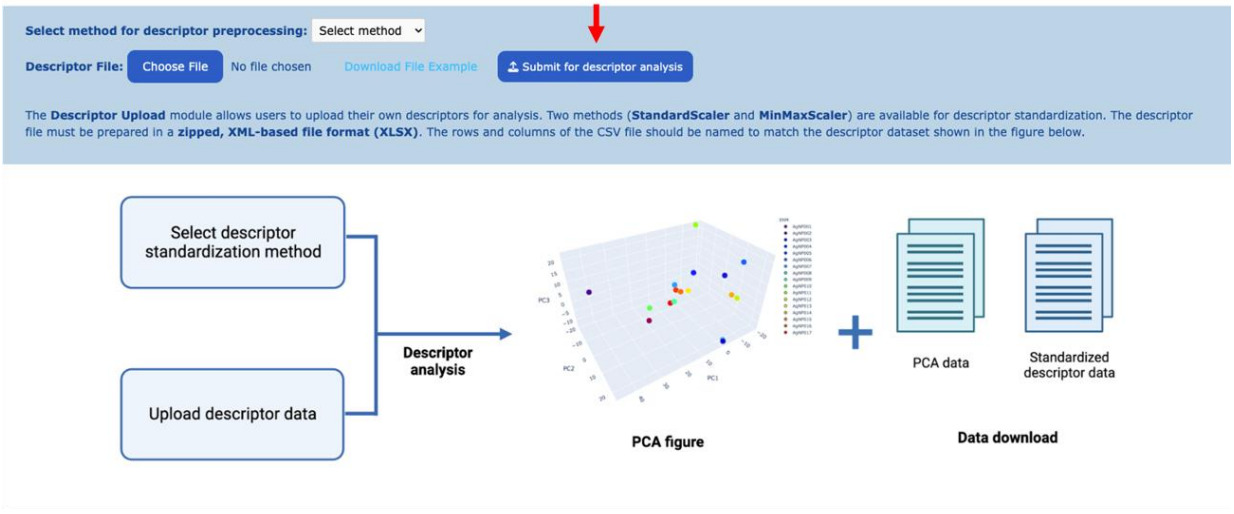


Figure 4.4 Nanodescriptors analysis through the Descriptor Upload module

Figure 4.5 Nanodescriptors analysis results from the Descriptor Upload approach

5. Model Toolkit

The Model toolkit has two modules: NanoPredictor and AutoNanoML for new NMs prediction through ML approach. The NanoPredictor module maintains series of pre-developed ML models for different prediction tasks. For example, the NanoPredictor interface of the PLSR model developed for NMs with assay 19 (NanoAID-19) and assay 20 (NanoAID-20) data is shown in **Figure 5.1**. It provides the model description, model-related literature, and an interactive scatter plot chart displaying the correlations between experimental and predicted values of the NMs used in the modeling. The interface allows for downloading the model in pickle (pkl) format, as well as the modeling datasets, including the nanodescriptor data and the

assay data in XLSX format. Users can locally prepare a nanodescriptor dataset of new NMs in XLSX format and submit it for prediction through the interface (**red arrow in Figure 5.1**). The interface will employ the pre-developed model for prediction and offer downloadable prediction results for users to evaluate these new NMs (**purple arrow in Figure 5.1**). Moreover, a dropdown menu is added to the module for switching between NanoPredictor interfaces with different pre-developed models, making it easy for users to perform various endpoint prediction tasks (**red box in Figure 5.1**).

The AutoNanoML module allow users to develop ML models through ViNAS-Pro. Two ML algorithms, linear regression (LR) and partial least squares regression (PLSR), are introduced for modeling in the module. For example, the initial AutoNanoML interface for developing PLSR model is shown in **Figure 5.2**. The modeling process can be divided into three steps: (1) uploading the descriptor and endpoint datasets in XLSX format; (2) choosing a descriptor standardization method, either StandardScaler or MinMaxScaler; and (3) selecting a cross-validation method among 3-Fold, 5-Fold, 10-Fold, or Leave-One-Out to develop the optimal ML model. After submitting for modeling, the AutoNanoML interface will update with new sections for model analysis and prediction (**Figure 5.3**). Users can visualize the model results through an interactive scatter plot chart that illustrates the correlations between experimental and predicted values of the NMs involved in the modeling. In addition, they can analyze the nanodescriptors by exploring an interactive pie chart that illustrates the contributions of the top-k descriptors derived from the modeling outcomes. The interface displays the optimal number of components for developing the best PLSR model, which is obtained from the cross-validation procedure. The R^2 and RMSE are two key metrics for users to assess the model performance. The model outcomes mentioned above are downloadable, including the model in pickle (pkl) format, the scatter plot chart data, and the descriptor contribution data. The updated interface also enables users to upload a nanodescriptor dataset of new NMs in XLSX format for prediction using their developed model (**red arrow and purple arrow in Figure 5.3**). The LR interface provides a workflow similar to PLSR, allowing users to develop LR models.

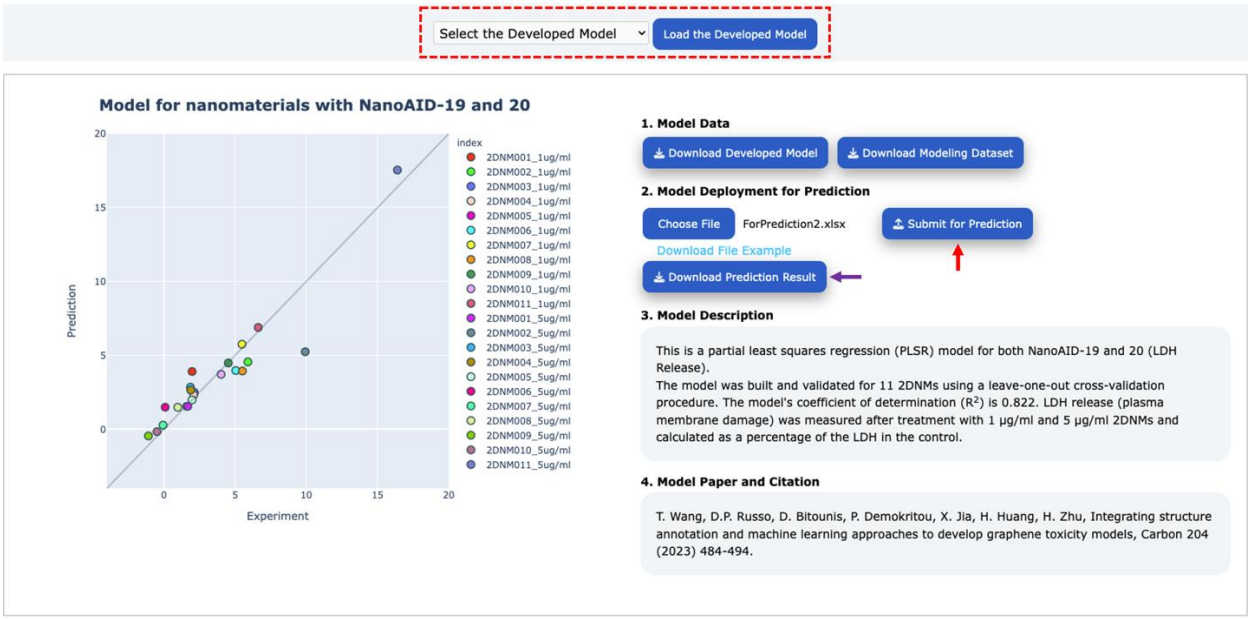


Figure 5.1 The NanoPredictor interface in the Model toolkit

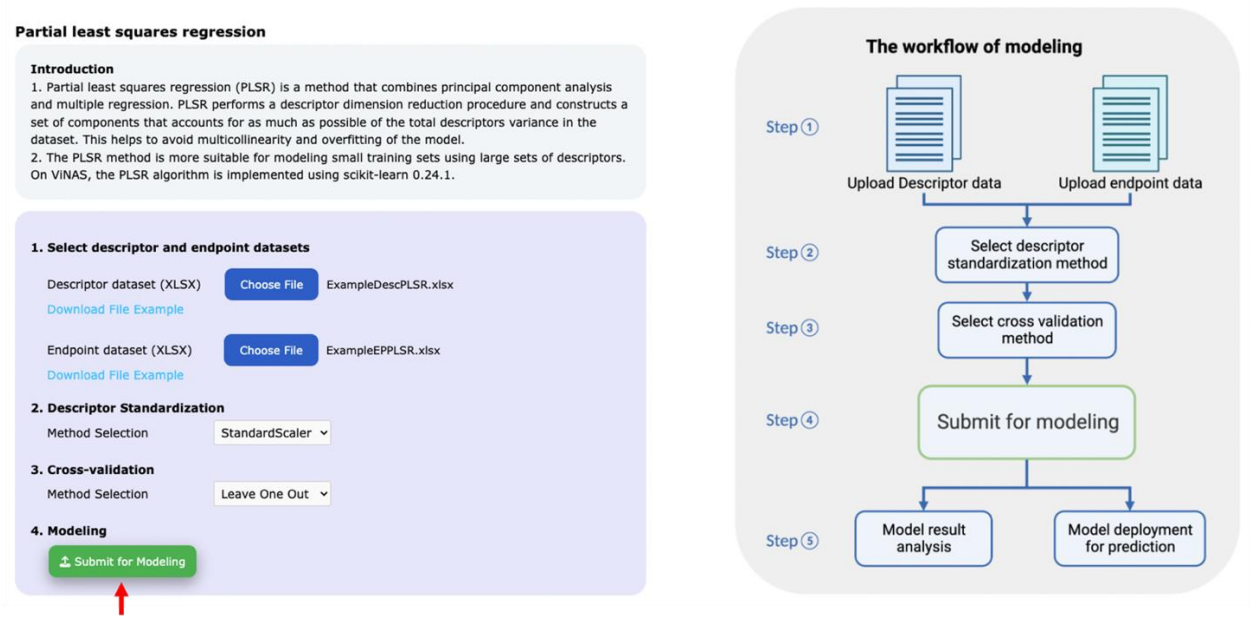


Figure 5.2 The initial AutoNanoML interface in the Model toolkit

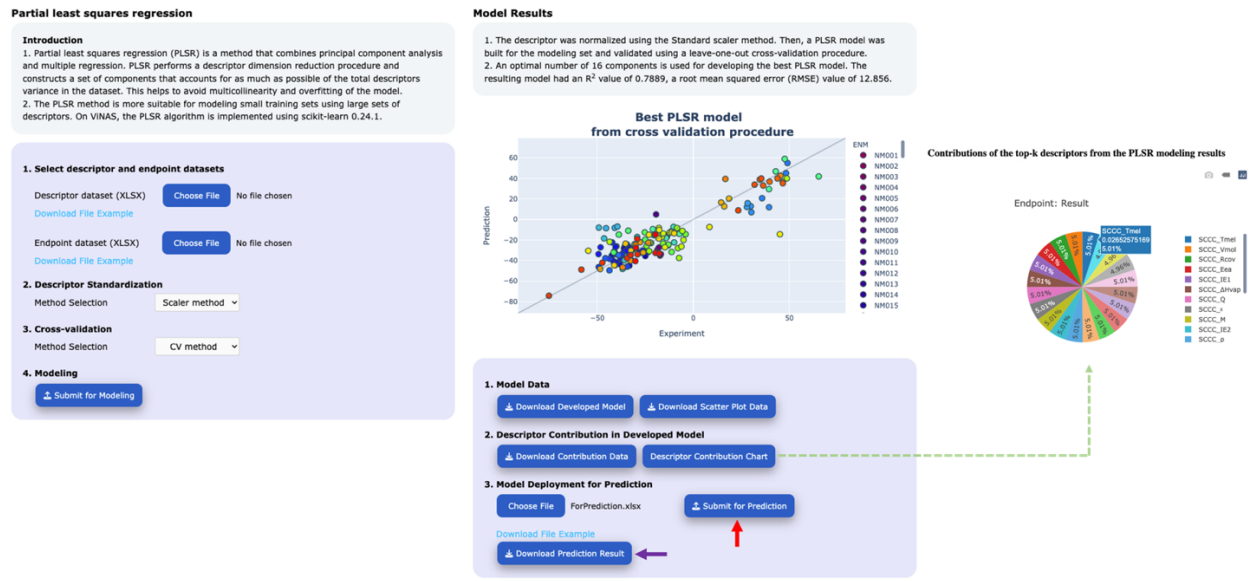


Figure 5.3 The updated AutoNanoML interface after ML modeling

6. The virtual Nanomaterials Library

To facilitate experimental research and reduce the time and cost associated with evaluating new NMs, a virtual NMs library is constructed and integrated into ViNAS-Pro, consisting of diverse nanostructures along with predictions of their properties, bioactivities, and toxicities. The library currently specializes in the development of virtual NMs for two material types, including platinum nanoparticles (PtNPs) and two-dimensional nanomaterials (2DNMs). For example, a total of 120 virtual PtNPs are developed based on the structural features of 12 experimentally synthesized PtNPs. The representative virtual PtNPs show structural diversity in size, types of surface ligands, and the number of surface ligands. The Library Analysis interface provides data analysis of virtual NMs in the library, categorized by different material types (**Figure 6.1**). The interface displays the size distribution chart, along with the 2D and 3D space charts of virtual NMs. It also provides an interactive table that lists the available virtual NMs in the library. The predictions for the virtual PtNPs using the pre-developed ML models are available and can be downloaded in batches through the Endpoint Profile interface (**Figure 6.2**). User can access detailed information and download relevant data for a specific virtual PtNP from the record page of the virtual NM by clicking on either the interactive table in the Library Analysis interface or the interactive table in the Endpoint Profile interface (**Figure 6.1 and 6.2**). The construction of virtual 2DNMs in the library is similar to that of the virtual PtNPs.

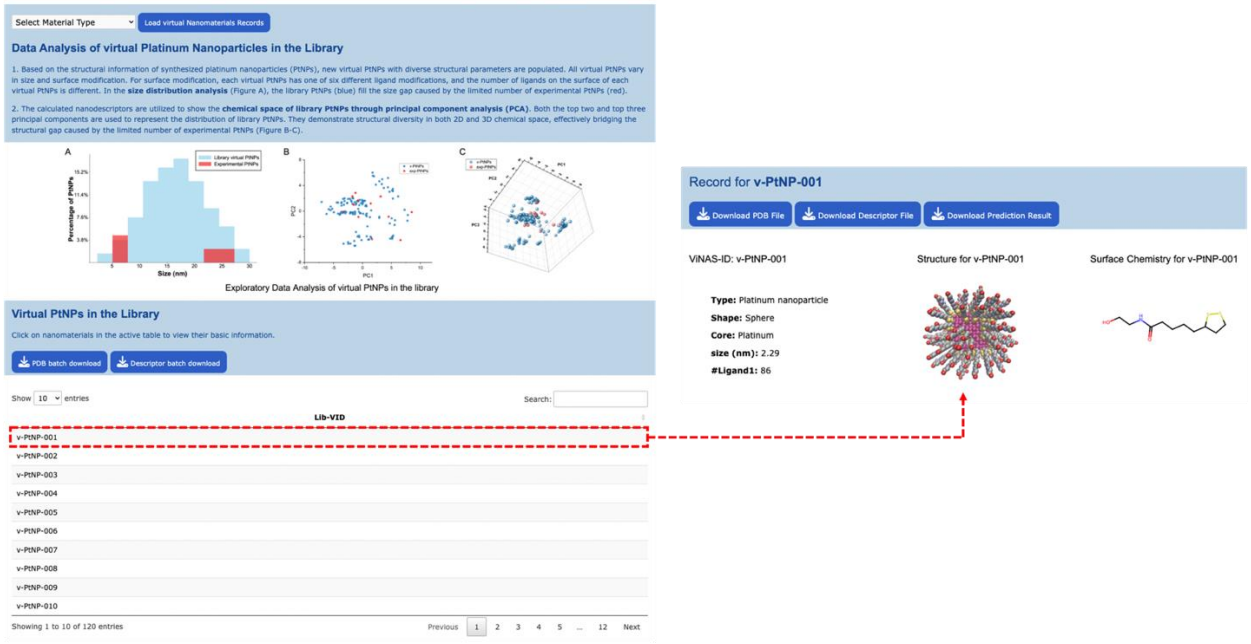


Figure 6.1 Exploring virtual nanomaterials through the Library Analysis interface

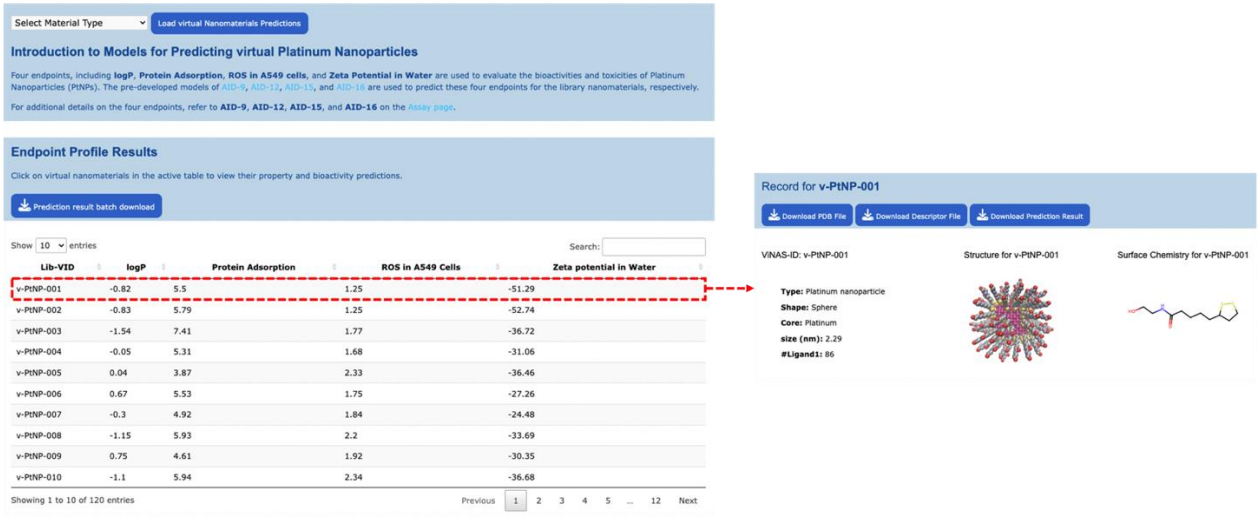


Figure 6.2 Exploring virtual nanomaterials through the Endpoint Profile interface

7. Data Deposit and Calculation Services

To facilitate data sharing within the nanoscience community, ViNAS-Pro provides a Data Deposit interface for users to deposit data into ViNAS-Pro databases (**Figure 7.1**). Depositors are encouraged to provide material information (e.g., material type and name), along with a corresponding reference (if available). They can then deposit nanostructure data in PDB format, nanodescriptor data, and assay results in CSV/XLSX format. After the in-house data cleaning

and validation, the uploaded data will be integrated into the ViNAS-Pro databases. Once the deposition is successful, depositors will receive a confirmation email.

ViNAS-Pro provides a nanostructure construction and nanodescriptor calculation service for new NMs (**Figure 7.2**). Users can request the service through the Calculation Service interface by providing basic information about the NM, including its name, shape, and size. It is preferable to include additional information, such as a corresponding reference, for the calculation service. When requesting nanodescriptor calculation, users are encouraged to provide corresponding nanostructures in PDB format. After the in-house calculation, users will receive a calculation result by email, which can be used for modeling and other nanoinformatics tasks.

For the two services mentioned above, users are required to provide an email address in the content/request details to receive the results of the service. This is an effective and secure method for protecting users' privacy. ViNAS-Pro ensures the protection of users' information and ownership of the data.

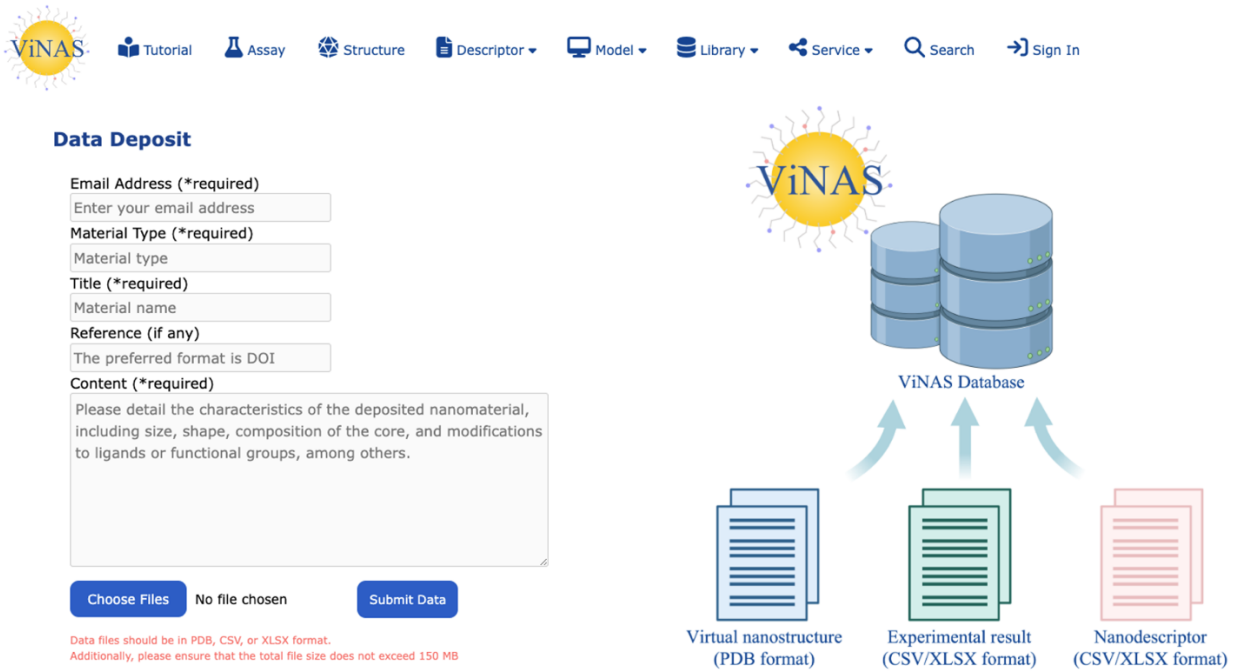


Figure 7.1 The Data Deposit interface

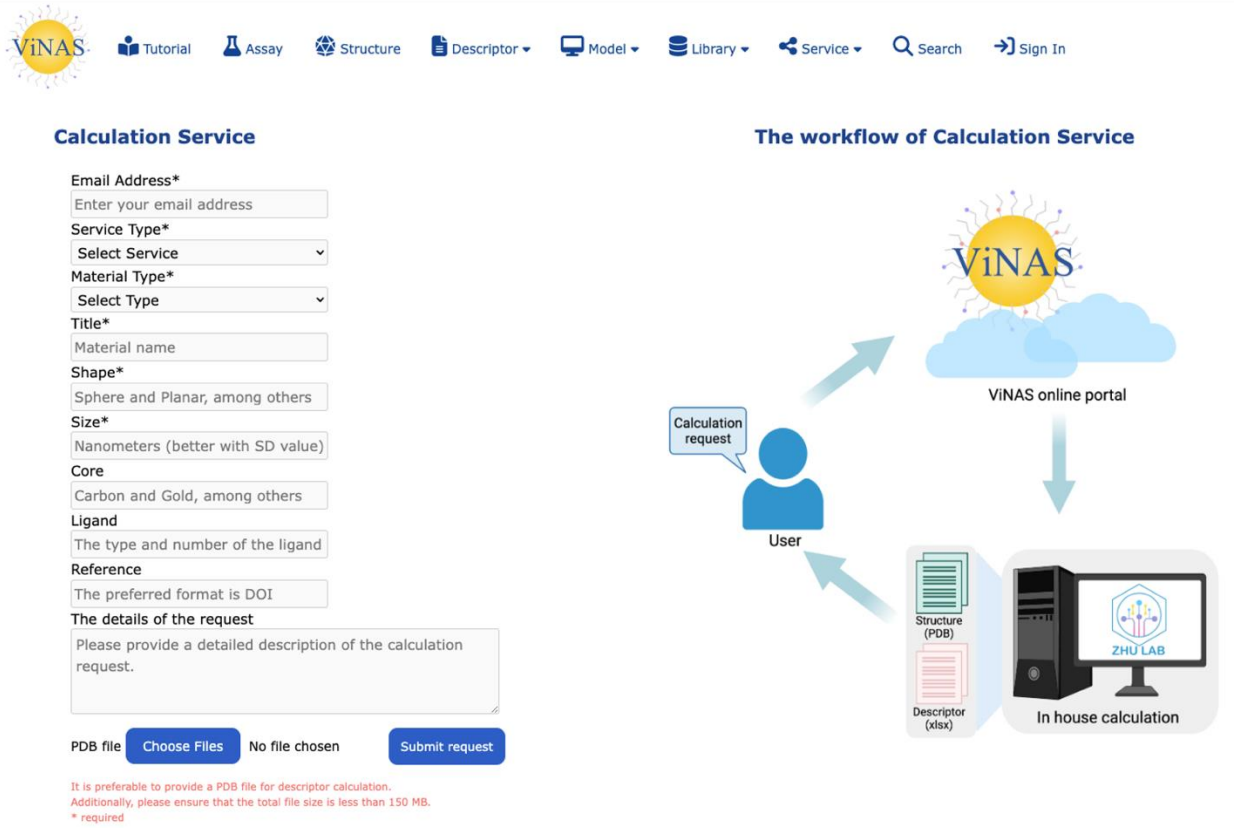


Figure 7.2 The Calculation Service interface

8. Search ViNAS-Pro databases

ViNAS-Pro allows users to search NM entries by different NM attributes, including ViNAS-ID (VID), NM type, shape, core type, and/or size (**Figure 8.1**). Users can focus on a single search criterion or combine multiple search criteria to find specific NM records. For example, users can search for NMs with ‘Sphere’ as the shape attribute, while keeping other search criteria unchanged. This results in a total of 583 NM entries that meet the search criteria, available for further browsing (**Figure 8.2**). Clicking on the ViNAS-ID in the results table will open the NM record page, displaying the NM's attributes, structure, and assay testing results. Moreover, a combination of search criteria, such as NM type and shape, can provide even more specific results. For example, users can search for NMs with 'Gold nanoparticle' as the NM type and 'Sphere' as the NM shape, while keeping other search criteria unchanged. This yields a total of 400 NM entries that meet the search criteria, available for further browsing (**Figure 8.3**). If all search criteria remain unchanged, the entire ViNAS-Pro database will be returned.

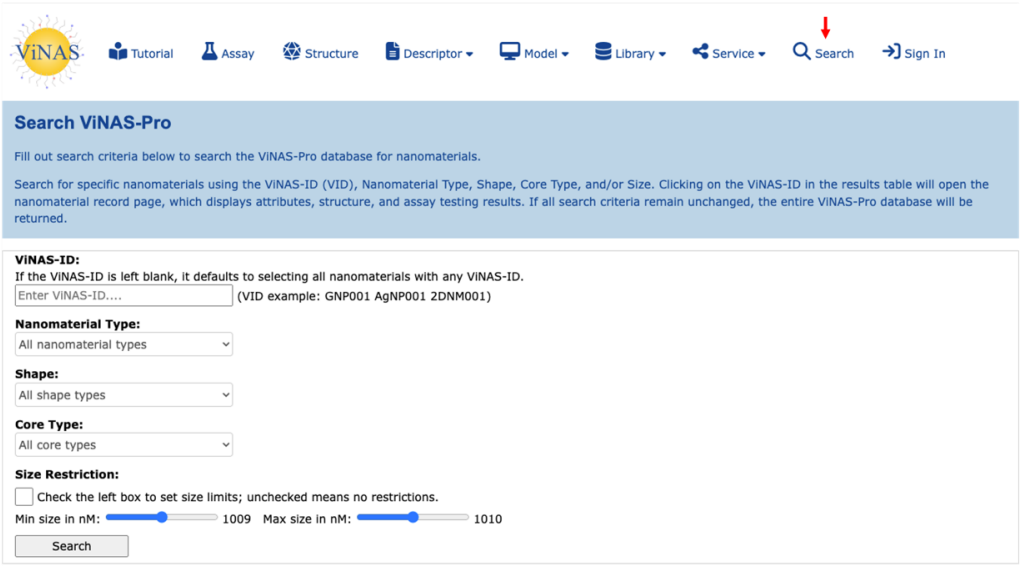


Figure 8.1 The Search interface

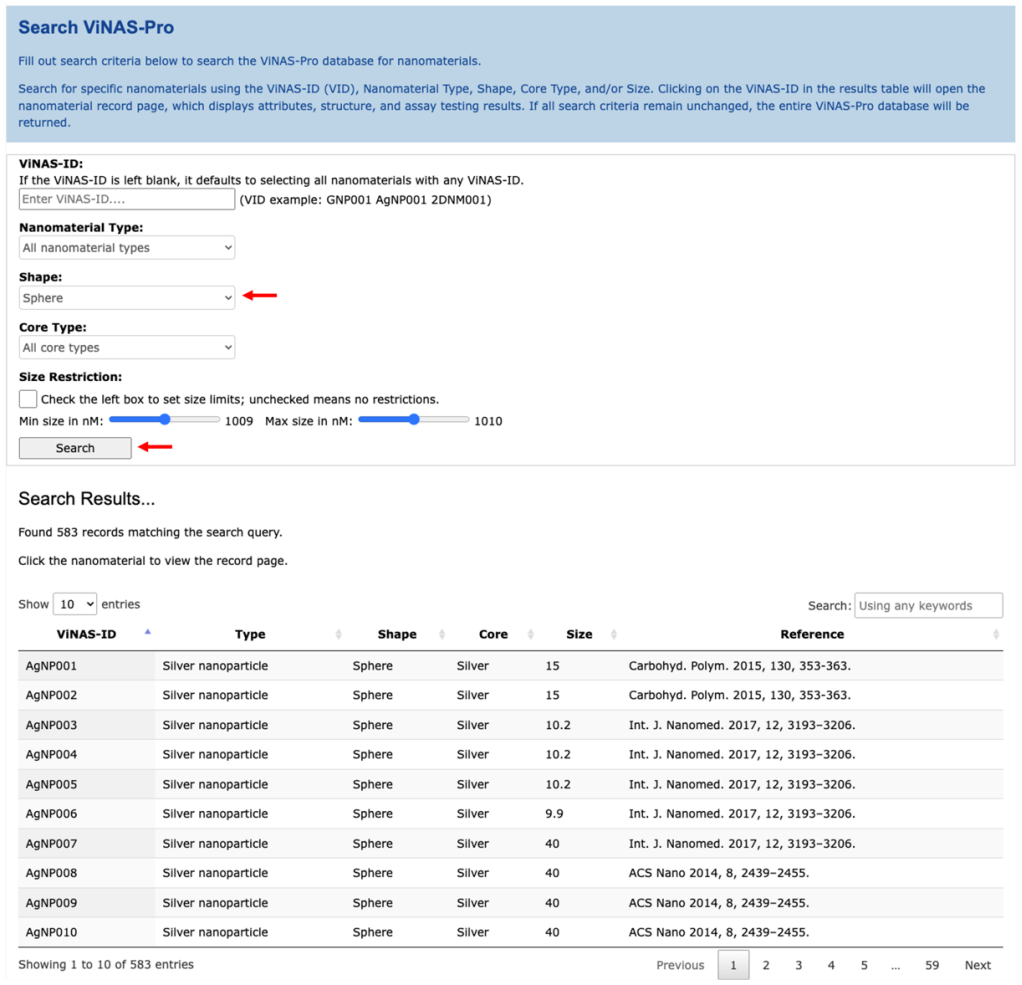


Figure 8.2 Search results for NMs with ‘Sphere’ as the shape attribute through the Search interface.

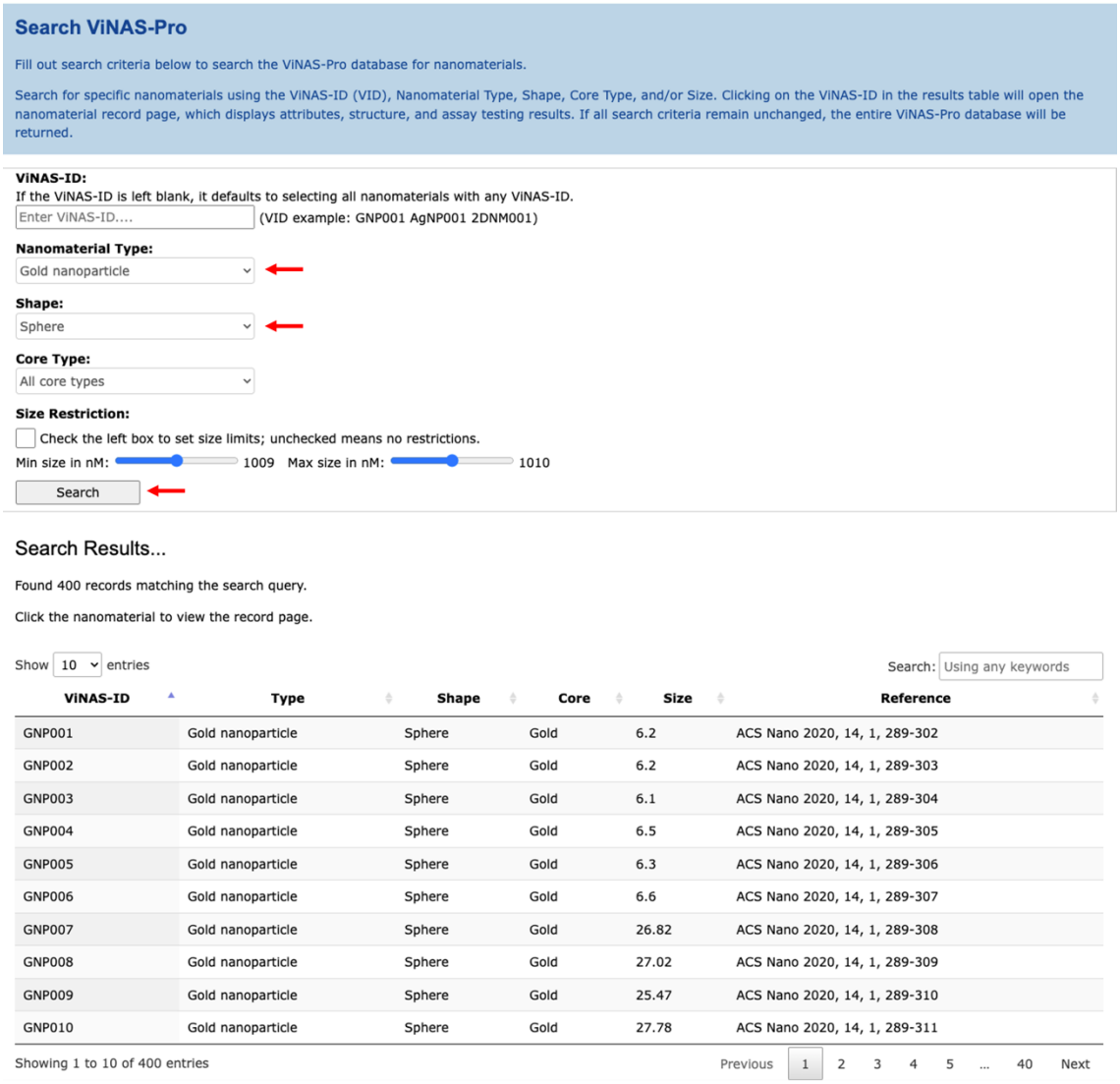


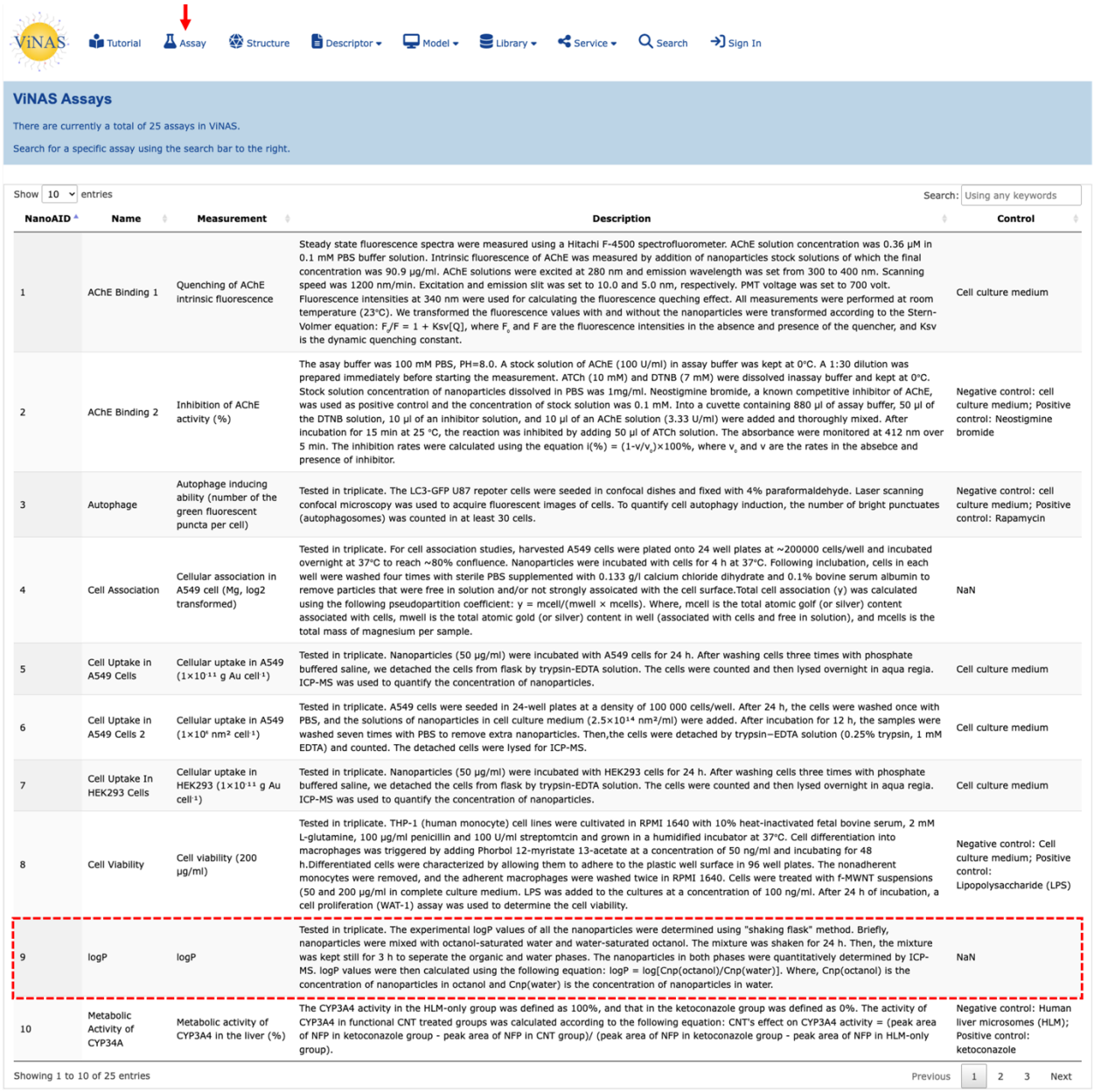
Figure 8.3 Search results for NMs with 'Gold nanoparticle' as the NM type and 'Sphere' as the NM shape through the Search interface.

9. Case Study: Regular Machine Learning Modeling Process

9.1 Data preparation

As mentioned in the Introduction, ViNAS-Pro provides access to a total of 25 assay data, covering 13 NM types. Users can select nanodescriptor and assay data for specific nanomaterials (NMs) from the Assay database, based on a particular assay, for modeling purposes. In this case study, we will select assay nine (NanoAID-9), which is the logP assay, for ML modeling (**Figure 9.1.1**). After clicking on assay nine from the table of navigation page, users will be directed to the record page of NanoAID-9, where a general introduction of NanoAID-9 is shown. The assay

data and descriptor data of NMs for NanoAID-9 can be downloaded, and the two datasets of 123 gold nanoparticles (GNP) will be used for ML modeling (**Figure 9.1.2 and 9.1.3**).



NanoAID	Name	Measurement	Description	Control
1	AChE Binding 1	Quenching of AChE intrinsic fluorescence	Steady state fluorescence spectra were measured using a Hitachi F-4500 spectrofluorometer. AChE solution concentration was 0.36 μM in 0.1 mM PBS buffer solution. Intrinsic fluorescence of AChE was measured by addition of nanoparticles stock solutions of which the final concentration was 90.9 μg/ml. AChE solutions were excited at 280 nm and emission wavelength was set from 300 to 400 nm. Scanning speed was 1200 nm/min. Excitation and emission slit was set to 10.0 and 5.0 nm, respectively. PMT voltage was set to 700 volt. Fluorescence intensities at 340 nm were used for calculating the fluorescence quecning effect. All measurements were performed at room temperature (23°C). We transformed the fluorescence values with and without the nanoparticles were transformed according to the Stern-Volmer equation: $F_0/F = 1 + Ksv[Q]$, where F_0 and F are the fluorescence intensities in the absence and presence of the quencher, and Ksv is the dynamic quenching constant.	Cell culture medium
2	AChE Binding 2	Inhibition of AChE activity (%)	The asay buffer was 100 mM PBS, PH=8.0. A stock solution of AChE (100 U/ml) in assay buffer was kept at 0°C. A 1:30 dilution was prepared immediately before starting the measurement. ATCh (10 mM) and DTNB (7 mM) were dissolved inassay buffer and kept at 0°C. Stock solution concentration of nanoparticles dissolved in PBS was 1mg/ml. Neostigmine bromide, a known competitive inhibitor of AChE, was used as positive control and the concentration of stock solution was 0.1 mM. Into a cuvette containing 880 μl of assay buffer, 50 μl of the DTNB solution, 10 μl of an inhibitor solution, and 10 μl of an AChE solution (3.33 U/ml) were added and thoroughly mixed. After incubation for 15 min at 25 °C, the reaction was inhibited by adding 50 μl of ATCh solution. The absorbance were monitored at 412 nm over 5 min. The inhibition rates were calculated using the equation $I(\%) = (1-v/v_0)\times100\%$, where v_0 and v are the rates in the absebce and presence of inhibitor.	Negative control: cell culture medium; Positive control: Neostigmine bromide
3	Autophage	Autophage inducing ability (number of the green fluorescent puncta per cell)	Tested in triplicate. The LC3-GFP U87 repoter cells were seeded in confocal dishes and fixed with 4% paraformaldehyde. Laser scanning confocal microscopy was used to acquire fluorescent images of cells. To quantify cell autophagy induction, the number of bright punctuates (autophagosomes) was counted in at least 30 cells.	Negative control: cell culture medium; Positive control: Rapamycin
4	Cell Association	Cellular association in A549 cell (Mg, log2 transformed)	Tested in triplicate. For cell association studies, harvested A549 cells were plated onto 24 well plates at ~200000 cells/well and incubated overnight at 37°C to reach ~80% confluence. Nanoparticles were incubated with cells for 4 h at 37°C. Following inclubation, cells in each well were washed four times with sterile PBS supplemented with 0.133 g/l calcium chloride dihydrate and 0.1% bovine serum albumin to remove particles that were free in solution and/or not strongly assoicated with the cell surface.Total cell association (y) was calculated using the following pseudopartition coefficient: y = mcell/(mwell × mcells). Where, mcell is the total atomic golf (or silver) content associated with cells, mwell is the total atomic gold (or silver) content in well (associated with cells and free in solution), and mcells is the total mass of magnesium per sample.	NaN
5	Cell Uptake in A549 Cells	Cellular uptake in A549 (1×10^{-11} g Au cell^{-1})	Tested in triplicate. Nanoparticles (50 μg/ml) were incubated with A549 cells for 24 h. After washing cells three times with phosphate buffered saline, we detached the cells from flask by trypsin-EDTA solution. The cells were counted and then lysed overnight in aqua regia. ICP-MS was used to quantify the concentration of nanoparticles.	Cell culture medium
6	Cell Uptake in A549 Cells 2	Cellular uptake in A549 (1×10^{6} nm^2 cell^{-1})	Tested in triplicate. A549 cells were seeded in 24-well plates at a density of 100 000 cells/well. After 24 h, the cells were washed once with PBS, and the solutions of nanoparticles in cell culture medium (2.5×10^{14} nm^2/ml) were added. After incubation for 12 h, the samples were washed seven times with PBS to remove extra nanoparticles. Then,the cells were detached by trypsin−EDTA solution (0.25% trypsin, 1 mM EDTA) and counted. The detached cells were lysed for ICP-MS.	Cell culture medium
7	Cell Uptake In HEK293 Cells	Cellular uptake in HEK293 (1×10^{-11} g Au cell^{-1})	Tested in triplicate. Nanoparticles (50 μg/ml) were incubated with HEK293 cells for 24 h. After washing cells three times with phosphate buffered saline, we detached the cells from flask by trypsin-EDTA solution. The cells were counted and then lysed overnight in aqua regia. ICP-MS was used to quantify the concentration of nanoparticles.	Cell culture medium
8	Cell Viability	Cell viability (200 μg/ml)	Tested in triplicate. THP-1 (human monocyte) cell lines were cultivated in RPMI 1640 with 10% heat-inactivated fetal bovine serum, 2 mM L-glutamine, 100 μg/ml penicillin and 100 U/ml streptomtcin and grown in a humidified incubator at 37°C. Cell differentiation into macrophages was triggered by adding Phorbol 12-myristate 13-acetate at a concentration of 50 ng/ml and incubating for 48 h.Differentiated cells were characterized by allowing them to adhere to the plastic well surface in 96 well plates. The nonadherent monocytes were removed, and the adherent macrophages were washed twice in RPMI 1640. Cells were treated with f-MWNT suspensions (50 and 200 μg/ml in complete culture medium. LPS was added to the cultures at a concentration of 100 ng/ml. After 24 h of incubation, a cell proliferation (WAT-1) assay was used to determine the cell viability.	Negative control: Cell culture medium; Positive control: Lipopolysaccharide (LPS)
9	logP	logP	Tested in triplicate. The experimental logP values of all the nanoparticles were determined using "shaking flask" method. Briefly, nanoparticles were mixed with octanol-saturated water and water-saturated octanol. The mixture was shaken for 24 h. Then, the mixture was kept still for 3 h to seperate the organic and water phases. The nanoparticles in both phases were quantitatively determined by ICP-MS. logP values were then calculated using the following equation: logP = log[Cnp(octanol)/Cnp(water)]. Where, Cnp(octanol) is the concentration of nanoparticles in octanol and Cnp(water) is the concentration of nanoparticles in water.	NaN
10	Metabolic Activity of CYP34A	Metabolic activity of CYP3A4 in the liver (%)	The CYP3A4 activity in the HLM-only group was defined as 100%, and that in the ketoconazole group was defined as 0%. The activity of CYP3A4 in functional CNT treated groups was calculated according to the following equation: CNT's effect on CYP3A4 activity = (peak area of NFP in ketoconazole group - peak area of NFP in CNT group)/ (peak area of NFP in ketoconazole group - peak area of NFP in HLM-only group).	Negative control: Human liver microsomes (HLM); Positive control: ketoconazole

Figure 9.1.1 Selecting an assay through the Assay database

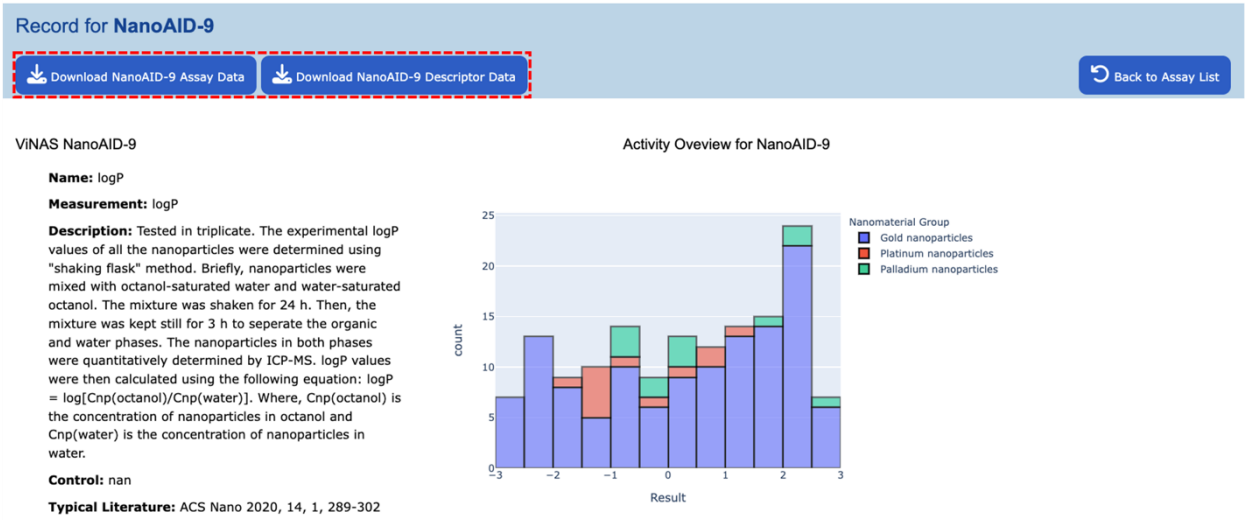


Figure 9.1.2 Downloading nanodescriptor data and assay data for NanoAID-9 associated NMs

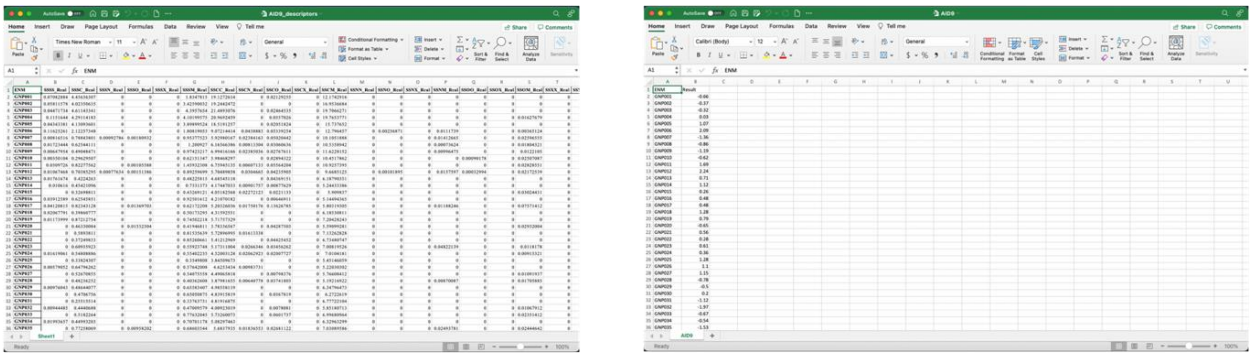


Figure 9.1.3 Nanodescriptor data and assay data for NanoAID-9 associated NMs in XLSX format

9.2 Descriptor analysis

After downloading the datasets, users can first analyze the nanodescriptor data of NanoAID-9 associated NMs through the Descriptor toolkit. Two methods (Standard Scaler and MinMax Scaler) are available for harmonizing nanodescriptor values. Users can upload the descriptor dataset through the Descriptor Upload module. After selecting the method and submitting for analysis, user can obtain PCA result and a descriptor standardization dataset (**Figure 9.2.1**).

PCA applied to the descriptor data identifies the combination of attributes (principal components, or directions in the descriptors space) that account for the most variance in the descriptor data. Based on the PCA results, both the 2D and 3D chemical spaces of NMs in the

dataset is visualized, which can be used to analyze their structure diversity (**Figure 9.2.2**). Descriptor standardization is a technique often applied as part of descriptor preparation for machine learning. The goal of standardization is to change the values of numeric columns in the dataset to a common scale, without distorting differences in the ranges of values.

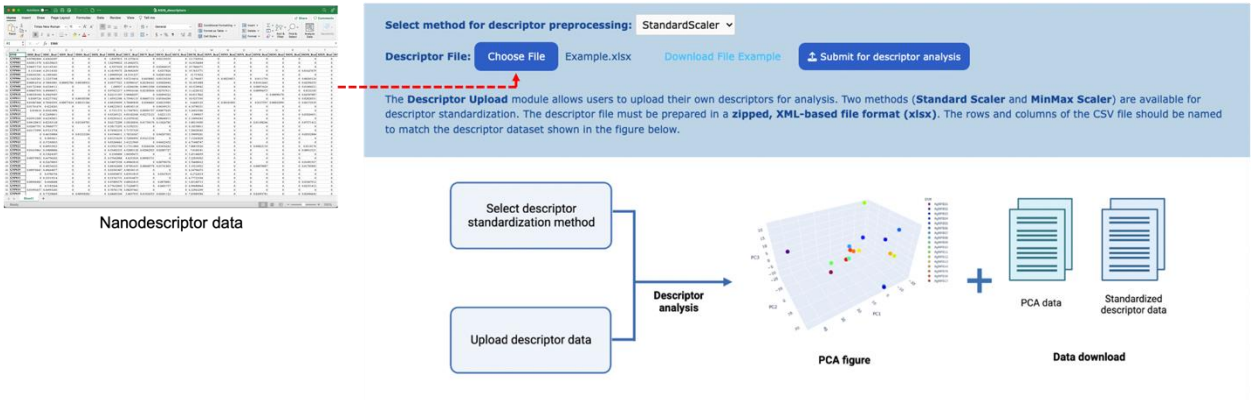


Figure 9.2.1 Descriptor analysis through the Descriptor Upload module

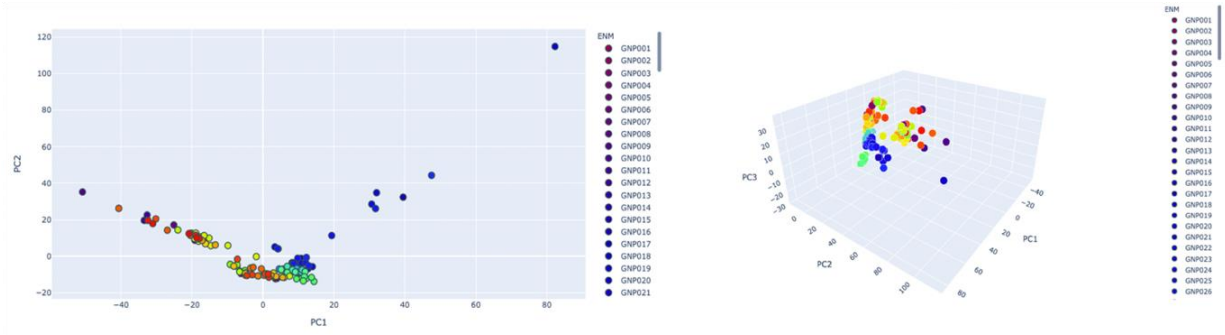


Figure 9.2.2 The two-dimensional and three-dimensional chemical spaces of NanoAID-9 associated NMs

9.3 Machine learning modeling

The Model toolkit allows users to predict NM properties and bioactivity using pre-developed models or by developing their own ML models through parameter tuning. In this case study, we will develop a LR model using the assay data and descriptor data of NMs associated with NanoAID-9 through the AutoNanoML module. In the LR page of AutoNanoML, users can develop customized LR models by three steps: (1) uploading descriptor dataset and assay dataset in XLSX format, (2) selecting a method for descriptor standardization, and (3) selecting a method for the cross-validation procedure (**Figure 9.3.1**). After clicking on the "Submit for Modeling" button, the process of model development is automated and streamlined for the user, and the resulting models and corresponding prediction results are available for browsing and

downloading (**Figure 9.3.2**). The development of a PLSR model is similar to that of an LR model. Users can explore the PLSR page of AutoNanoML for PLSR modeling.

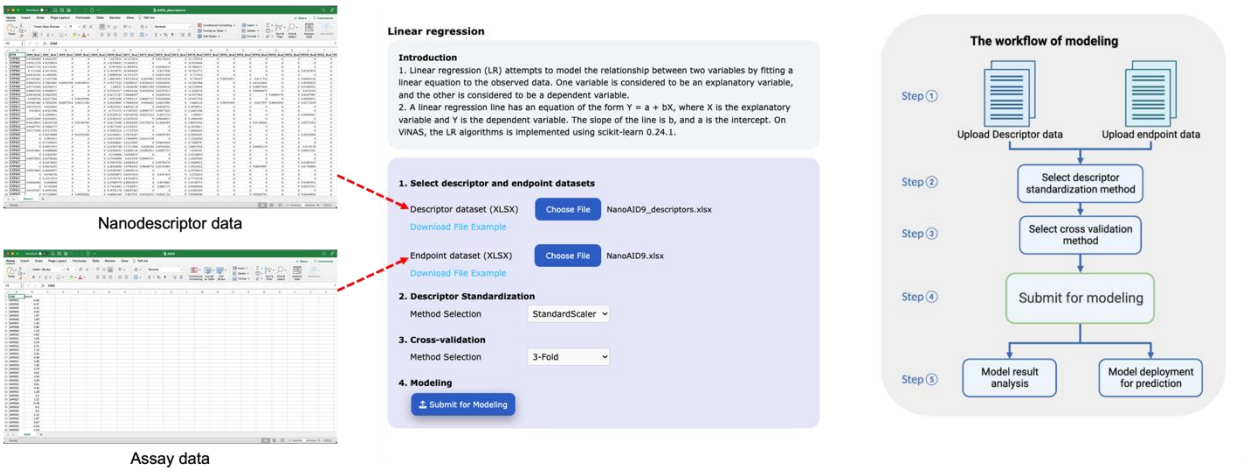


Figure 9.3.1 Machine learning modeling through the AutoNanoML module

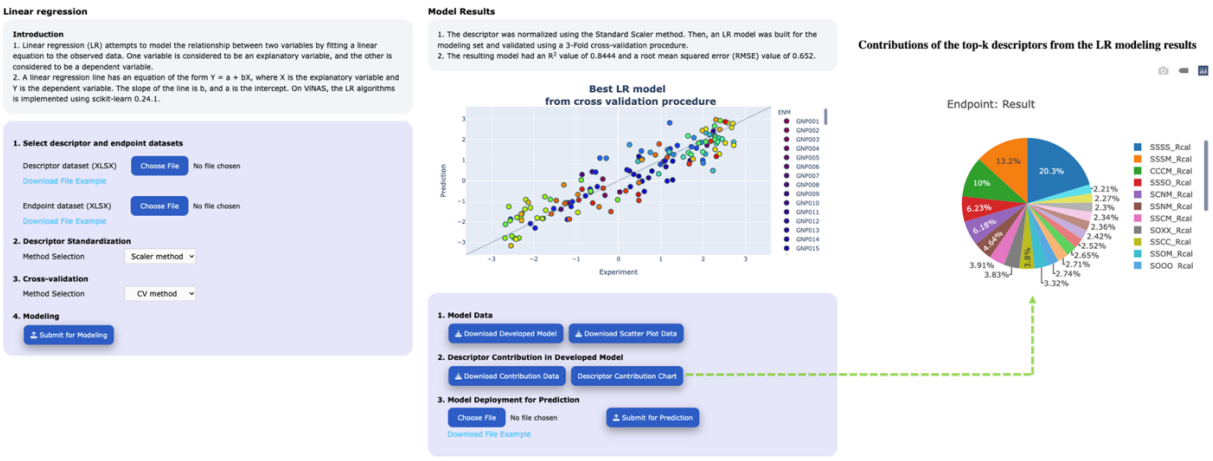


Figure 9.3.2 The developed model using the nanodescriptor data and assay data of NanoAID-9 associated NMs

9.4 Prediction new nanomaterials using the developed model

In section 9.1-9.3, we introduced a case to develop a LR model using the NMs associated with NanoAID-9. The subsequent step involves utilizing the developed model to predict the logP values of new NMs. Users need to create a nanodescriptor dataset of new NMs in XLSX format for this purpose. There are several ways to obtain specific NMs' nanodescriptor data for prediction. A regular approach is to explore and retrieve data from the Structure database. Moreover, as described in section 7, users can also request a nanodescriptor calculation service

through the Calculation Service interface by providing basic structure information about the new NMs.

In this tutorial, we will focus on GNPs in the Structure database, since the LR model is developed using GNPs. After clicking on “Gold Nanoparticles” on the primary navigation page of the Structure database, users will be directed to the secondary navigation page for GNPs, where a total of 414 GNPs nanodescriptor data can be batch downloaded in XLSX format (**Figure 9.4.1 and 9.4.2**). In section 9.1-9.3, we use the nanodescriptor dataset and assay dataset of 123 GNPs for modeling, while the other 291 GNPs possess nanodescriptor data but lack assay data of NanoAID-9 (**Figure 9.4.3**). The nanodescriptor data for all 414 GNPs can be obtained from the secondary navigation page by clicking on 'Descriptor batch download' (**Figure 9.4.2**). Subsequently, we will use the nanodescriptor data from the GNPs without NanoAID-9 results to create a prediction dataset in XLSX format.

The nanodescriptor dataset for predicting new NMs will be uploaded to the developed LR model on the LR page, as mentioned in section 9.3 (**Figure 9.4.4**). After clicking the 'Submit for Prediction' button, the model will perform the prediction, and the prediction results can be downloaded and analyzed from the interface (**Figure 9.4.5**).

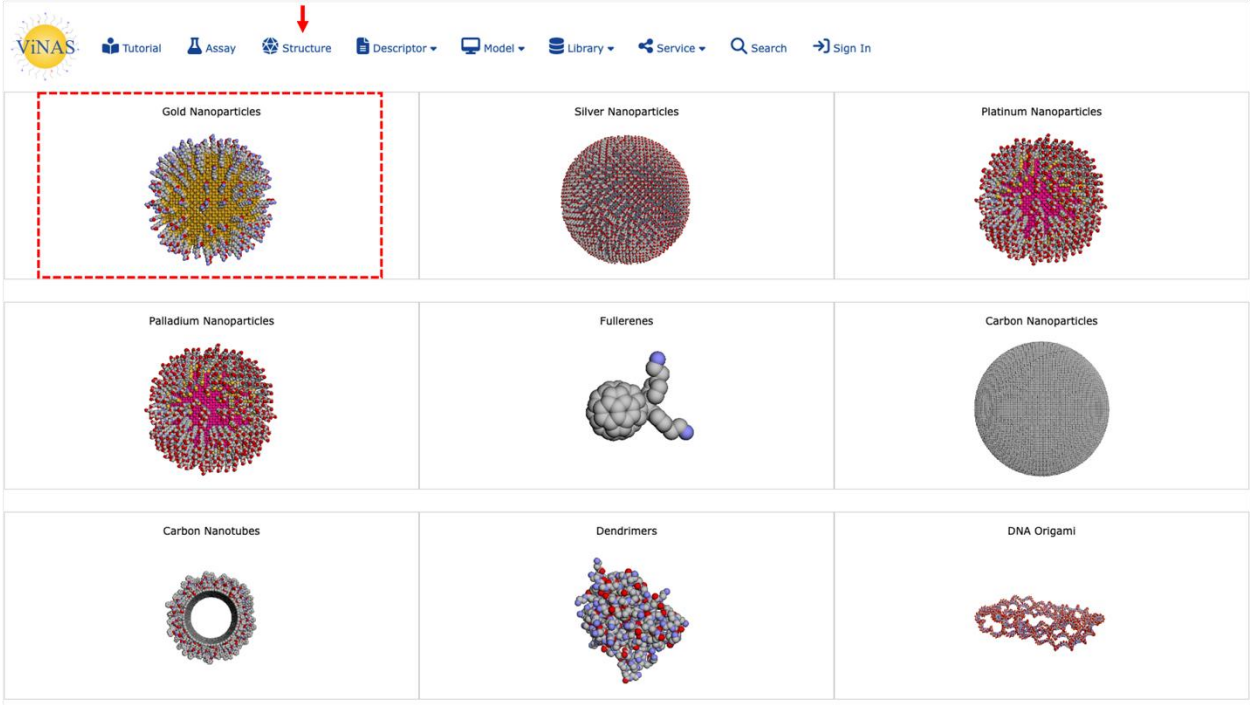


Figure 9.4.1 Selecting specific nanomaterial type from the primary navigation page of the Structure database

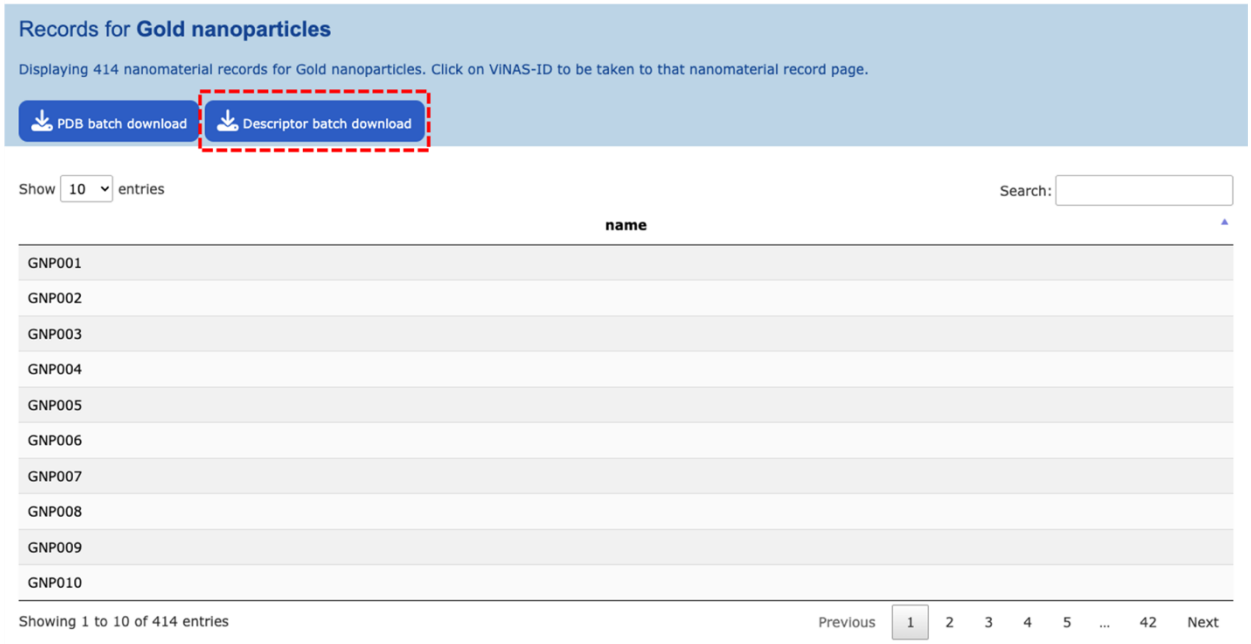


Figure 9.4.2 Batch download of GNPs nanodescriptor data from the secondary navigation page of the Structure database

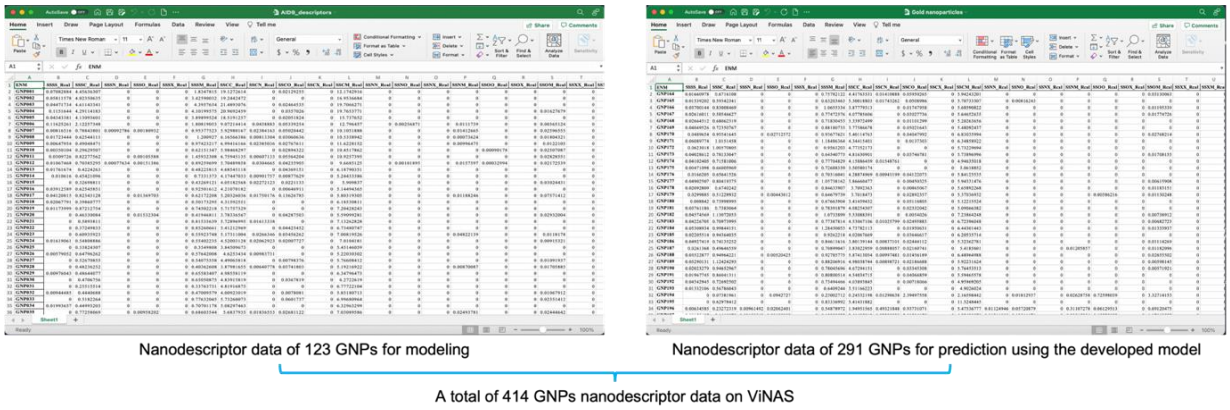


Figure 9.4.3 GNPs nanodescriptor data used for machine learning modeling and prediction

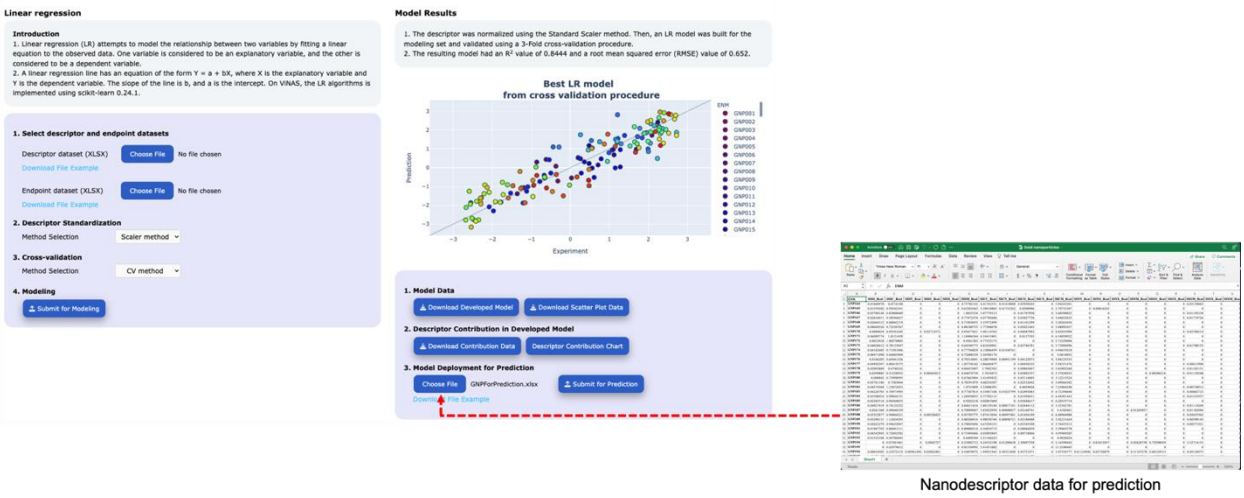


Figure 9.4.4 Uploading the nanodescriptor dataset of new nanomaterials for prediction

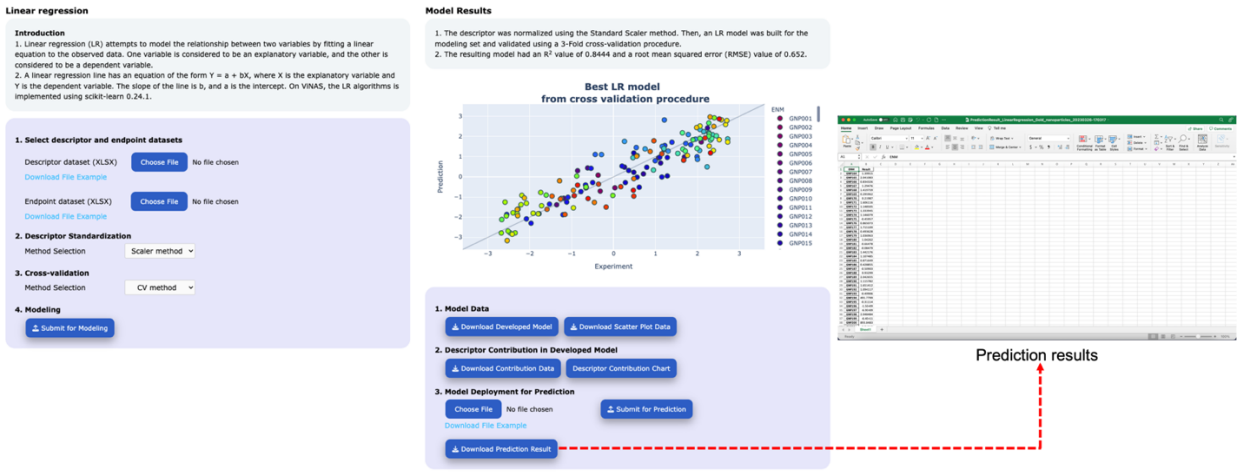


Figure 9.4.5 Downloading the prediction results from the AutoNanoML interface

10. Contact Us

We appreciate user feedback on ViNAS-Pro and aim to respond to all inquiries. Users can reach us by email at vinas.zhulab@gmail.com. We will continuously upgrade ViNAS-Pro to better serve the community.

11. About

The Zhu Lab uses cheminformatics algorithms, workflows, and other relevant computational tools to model chemical toxicity, ADME (Absorption, Distribution, Metabolism, and Excretion), and other biological activities. The resulting models will be used in the

regulatory chemical toxicity assessments and the CADD (Computer-Aided Drug Discovery) process. To learn more about our lab, please visit our website at https://www.zhuhlab.com/.

2023.12.29 Version