synthesized since the dawn of history is approximately (the ninth power of 10). However, it is almost impossible to examine the binding force with respect to all the astronomical numbers of compounds by experiments or simulations. Even in a case of examination of the binding force with respect to some compounds, the efficiency is low only by repeating trial and error as described in US9373059B and JP5946045B. Further, in the case of a descriptor (feature quantity) in the related art such as the fingerprint described in JP4564097B, the feature quantities of compounds are not necessarily similar even in a case where the compounds exhibit the same drug efficacy. Further, since the feature quantities did not accurately show the chemical properties of the target structure, the efficiency of search using the feature quantities was low.

[0006] As described above, in the related art, feature quantities do not accurately show the chemical properties of the target structures, and thus the efficiency of screening using the feature quantity and creation of a three-dimensional structure is low.

[0007] Embodiments of the present invention provide a feature quantity calculating method, a feature quantity calculating program, and a feature quantity calculating device which enable calculation of a feature quantity accurately showing chemical properties of a target structure. Further, embodiments of the present invention provide a screening method and a screening program which enable efficient screening of a pharmaceutical candidate compound using a feature quantity. Further, embodiments of the present invention provide a compound creating method which enable efficient creation of a three-dimensional structure of a pharmaceutical candidate compound using a feature quantity.

[0008] According to a first aspect of the present invention, there is provided a feature quantity calculating method including: a target structure designating step of designating a target structure formed of a plurality of unit structures having chemical properties; and a feature quantity calculating step of calculating a feature quantity obtained by quantifying, in a three-dimensional space, a degree of accumulation of one or more kinds of probes in a periphery of a three-dimensional structure of the target structure and calculating the feature quantity from the target structure using a generator formed through machine learning, in which the probe is a structure in which a plurality of points having a real electric charge and generating a van der Waals force are disposed to be separated from each other.

[0009] Since the chemical properties of target structures are exhibited as the result of an interaction between the target structure and one or more kinds of probes in the periphery thereof, the fact that target structures have a similarity in the degree of

accumulation of probes indicates that the chemical properties of the target structures are similar. That is, target structures having similar feature quantities calculated according to the first aspect exhibit similar chemical properties. Therefore, according to the first aspect, the feature quantity accurately showing the chemical properties of a target structure can be calculated.

[0010] In the first aspect, the "generator" (prediction model, learned model) used for calculating a feature quantity can be constructed by "a method of constructing a generator, including a learning processing step of inputting [a data set for learning using information (three-dimensional images and the like) related to the three-dimensional structures of compounds as teacher data and the feature quantities as explanatory variables for a plurality of compounds] to the generator to perform learning through machine learning". The generator may be constructed using a hierarchical neural network.

[0011] In the first aspect, the "point" constituting a probe is different from the mathematical "point" and may have a size (an actual atom, an actual ion, or the like). Further, a virtual point electric charge (one aspect of the "point"), a dipole, or the like may be used as a probe instead of an actual atom or the like. The probe can be selected according to a target compound (target structure).

[0012] Further, the feature quantity calculating method according to the first aspect can be performed by a computer realized using one or more of various processors such as a central processing unit (CPU) or a feature quantity calculating device, a feature quantity calculating system, or the like comprising such a computer. The same applies to the feature quantity calculating method, the screening method, and the compound creating method according to each of the following aspects.

[0013] In a first embodiment of the feature quantity calculating method , a compound is designated as the target structure in the target structure designating step, and a first feature quantity which is a feature quantity obtained by quantifying, in the three-dimensional space, a degree of accumulation of amino acids as the probes in the periphery of the three-dimensional structure of the compound is calculated using the generator in the feature quantity calculating step. In the first embodiment, the "probe", the "target structure", and the "plurality of unit structures" in the first aspect are respectively an amino acid, a compound, and a plurality of atoms. Further, the number of kinds of amino acids used for quantifying the degree of accumulation is not limited to one, and a peptide in which two or more kinds of amino acids are bound may be used.

[0014] Similar to the first aspect, since the drug efficacy of a compound (the binding force with respect to the target protein) is locally exhibited as the result of an interaction between the compound and each amino acid (one aspect of a probe), in a case where the degree of accumulation of amino acids is similar between compounds, the compounds have similar binding forces (with respect to the target proteins). That is, compounds having similar feature quantities (first feature quantities) according to the first embodiment exhibit similar drug efficacies. Therefore, according to the first embodiment, the feature quantity accurately showing the chemical properties of a compound can be calculated. Further, in the first embodiment, a compound such as a biological ligand which has a three-dimensional structure and whose binding to a target protein is known can be designated as a target structure.

[0015] In a second embodiment of the feature quantity calculating method, a pocket structure bound to a pocket that is an active site of a target protein is designated as the target structure in the target structure designating step, and a second feature quantity which is a feature quantity obtained by quantifying, in the three-dimensional space, a degree of accumulation of amino acids as the probes in the periphery of a three-dimensional structure of the pocket structure formed of a plurality of virtual spheres is calculated using the generator in the feature quantity calculating step. In the second embodiment, the "probe", the "target structure", and the "unit structure" in the first aspect are respectively an amino acid, a pocket structure, and a plurality of virtual spheres. The "active site" of the target protein indicates a site where the activity of the target protein is promoted or suppressed by binding a pocket structure, and the "virtual sphere" can be considered to have chemical properties such as the van der Waals radius and the electric charge.

[0016] Similar to the description of the first embodiment, since the pocket structures having similar feature quantities (second feature quantities) according to the second embodiment exhibit similar chemical properties, the feature quantity accurately showing the chemical properties of the pocket structure can be calculated according to the third aspect. Further, the pocket structure corresponds to a compound that is bound to a pocket of the target protein. Further, in the second embodiment, as the result of actual measurement on the three-dimensional structure of the target protein, simulation based on position information and the like of the pocket can be used for calculation of the second feature quantity. Further, the measuring techniques (for example, an X-ray crystal structure, a nuclear magnetic resonance (NMR) structure, and a cryo-transmission electron microscopy (TEM) structure) are not limited as long

as the three-dimensional structure of the target protein is a three-dimensional structure with a resolution that enables identification of each residue of an amino acid.

[0017] In a third embodiment of the feature quantity calculating method, a compound is designated as the target structure in the target structure designating step, and a third feature quantity which is a feature quantity obtained by quantifying, in the three-dimensional space, a degree of accumulation of the probes in the periphery of the three-dimensional structure of the compound which is the degree of accumulation using one or more selected from one or more kinds of nucleic acid bases, one or more kinds of lipid molecules, one or more kinds of monosaccharide molecules, water, or one or more kinds of ions formed of a plurality of atoms, as the probes is calculated using the generator in the feature quantity calculating step. In the third embodiment, the "probe", the "target structure", and the "plurality of unit structures" in the first aspect are respectively one or more kinds of nucleic acid bases (the kind, the number, and the combination thereof may be optional), a compound, and a plurality of atoms.

[0018] In embodiments of the present invention, deoxyribonucleic acid (DNA), ribonucleic acid (RNA), cell membranes, and polysaccharides, which are biopolymers (compounds) other than proteins can be treated as the targets of drugs. In the third embodiment, the method of calculating the feature quantities of these target compounds is defined, and the probe is not an amino acid but another substance (a building block of each target). Specifically, in a case where the targets are DNA, RNA, cell membranes, and polysaccharides, the probes are respectively set as one or more kinds of nucleic acid bases, one or more kinds of nucleic acid bases, one or more kinds of lipid molecules, and one or more kinds of monosaccharide molecules. Further, water and one or more kinds of ions formed of a plurality of atoms may be considered in a case of the quantification of the degree of accumulation using these as probes. In the same manner as in the above embodiments, since the drug efficacy of a compound (the binding force with respect to the target such as DNA) is locally exhibited as the result of an interaction between a compound and a nucleic acid base or the like (probe), in a case where the degree of accumulation of nucleic acid bases and the like is similar between compounds, the compounds have similar binding forces with respect to the targets. That is, compounds having similar feature quantities (third feature quantities) according to the third embodiment exhibit similar drug efficacies. Therefore, according to the third embodiment, the feature quantity accurately showing the chemical properties of a compound can be calculated.

[0019] In the above embodiments, an invariant conversion step of converting the feature quantity into an invariant with respect to rotation and translation of the

compound may be provided. By converting the feature quantity into an invariant, the feature quantity is easily handled and the data capacity can be reduced. The conversion of the feature quantity into an invariant can be performed by Fourier transform, angular integration of a correlation function, or the like.

[0020] According to a second aspect of the present invention, there is provided a feature quantity calculating program which causes a computer to execute the feature quantity calculating method according to the above embodiments. The "computer" in the second aspect can be realized using one or more of various processors such as a central processing unit (CPU). This "computer" may be a computer comprised in a feature quantity calculating device or a feature quantity calculating system. Further, a non-transitory recording medium on which a computer-readable code of the feature quantity calculating program according to the second aspect is recorded can also be exemplified as an aspect of the present invention.

[0021] According to a third aspect of the present invention, there is provided a feature quantity calculating device including: a target structure designation unit which designates a target structure formed of a plurality of unit structures having chemical properties; and a feature quantity calculation unit which calculates a feature quantity obtained by quantifying, in a three-dimensional space, a degree of accumulation of one or more kinds of probes in a periphery of a three-dimensional structure of the target structure and calculating the feature quantity from the target structure using a generator formed through machine learning, in which the probe is a structure in which a plurality of points having a real electric charge and generating a van der Waals force are disposed to be separated from each other. According to the third aspect, similarly to the first aspect, the feature quantity accurately showing the chemical properties of the target structure can be calculated. Further, the feature quantity calculating device according to the third aspect may perform the feature quantity calculating method according to embodiments of the present invention by reading the code of the feature quantity calculating program according to embodiments of the present invention using various processors such as a CPU.

[0022] According to a fourth aspect of the present invention, there is provided a screening method of extracting a target compound which is bound to a target protein from a plurality of compounds, the method including: a storing step of storing a three-dimensional structure of a compound formed of a plurality of atoms and the first feature quantity of the three-dimensional structure of the compound which is calculated using the feature quantity calculating method according to the first embodiment in association with each other for each of the plurality of compounds; a

feature quantity calculating step of calculating the first feature quantity of a ligand that is a compound whose binding to the target protein has been confirmed using the feature quantity calculating method according to the first embodiment; a similarity calculating step of calculating a similarity between the first feature quantity of the plurality of compounds and the first feature quantity of the ligand; and a compound extracting step of extracting the target compound from the plurality of compounds based on the similarity. As the description for the first embodiment, in a case where the ligand and the target compound have similar first feature quantities (the feature quantities three-dimensionally quantified), the drug efficacies of both the ligand and the target compound are similar. Therefore, according to the fourth aspect, a target compound having drug efficacy similar to that of the ligand can be extracted based on the first feature quantity so that screening of a pharmaceutical candidate compound can be efficiently performed. Further, a compound that has a similarity satisfying the target value (for example, greater than or equal to the threshold) may be extracted or a compound may be extracted in a descending order of the similarity in the compound extracting step.

[0023] According to a fifth aspect of the present invention, there is provided a screening method of extracting a target compound which is bound to a target protein from a plurality of compounds, the method including: a storing step of storing a three-dimensional structure of a compound formed of a plurality of atoms and the first feature quantity calculated using the feature quantity calculating method according to the first embodiment in association with each other for each of the plurality of compounds; a feature quantity calculating step of calculating the second feature quantity of the pocket structure of the target protein using the feature quantity calculating method according to the second embodiment; a similarity calculating step of calculating a similarity between the first feature quantity of the plurality of compounds and the second feature quantity of the pocket structure; and a compound extracting step of extracting the target compound from the plurality of compounds based on the similarity.

[0024] As the description for the second embodiment, in a case where the pocket structure and the target compound have similar second feature quantities, the chemical properties of both the pocket structure and the target compound are similar. Therefore, according to the fifth aspect, a target compound having chemical properties similar to those of the pocket structure is extracted so that screening of a pharmaceutical candidate compound can be efficiently performed. Since the pocket structure corresponds to the compound that is bound to the target protein, the feature

quantity (the second feature quantity) of the pocket structure can be compared with the feature quantity (the first feature quantity) of the compound, and the similarity can be calculated.

[0025] According to a sixth aspect of the present invention, there is provided a screening method of extracting a target compound which is bound to a target biopolymer other than a protein from a plurality of compounds, the method including: a storing step of storing a three-dimensional structure of a compound formed of a plurality of atoms and the third feature quantity of the three-dimensional structure of the compound which is calculated using the feature quantity calculating method according to the third embodiment in association with each other for each of the plurality of compounds; a feature quantity calculating step of calculating the third feature quantity of a binding compound that is a compound whose binding to the target biopolymer other than the protein has been confirmed using the feature quantity calculating method according to the third embodiment; a similarity calculating step of calculating a similarity between the third feature quantity of the plurality of compounds and the third feature quantity of the binding compound; and a compound extracting step of extracting the target compound from the plurality of compounds based on the similarity. As described above for the third embodiment, DNA or the like which is a target biopolymer other than a protein can be handled, and in a case where the target compound and the binding compound that is bound to the target biopolymer have similar third feature quantities, the drug efficacies of both the target compound and the binding compound are similar. Therefore, according to the sixth aspect, a target compound having drug efficacy similar to that of the binding compound is extracted based on the third feature quantity so that screening of a pharmaceutical candidate compound can be efficiently performed.

[0026]According to a seventh aspect of the present invention, there is provided a screening method including: an input step of inputting information related to three-dimensional structures of a plurality of compounds; a feature quantity calculating step of calculating the feature quantity of each of the plurality of compounds using the feature quantity calculating method according to any one of the above embodiments based on the information; a characteristic value calculating step of calculating a physical property value indicating a physical property and/or a toxicity value indicating toxicity of each of the plurality of compounds based on the information; and an extracting step of extracting compounds in which a similarity between the feature quantities satisfies a target value and the physical property value and/or the toxicity value satisfies a target value, from the plurality of compounds.

[0027] According to the feature quantities of the present invention (the feature quantities calculated by the feature quantity calculating method according to the above embodiments), even in a case where an existing compound has a problem in terms of physical properties and toxicity, the skeleton of the compound can be significantly changed by recreating a novel compound having the same binding force to the target protein or the like (that is, the feature quantities of the novel compound and the existing compound are the same as each other), and thus the possibility of overcoming the problem of the physical properties and the toxicity can be increased. Further, in order to achieve both "maintenance of the binding force" and "avoidance of the problems of physical properties and toxicity", it is effective to combine the feature quantities according to the present invention with indices showing the physical properties and the toxicity. The seventh aspect defines one aspect of the screening method based on such an idea, and compounds in which a similarity between the feature quantities satisfies the target value and the physical property value and/or the toxicity value satisfies the target value are extracted from the plurality of compounds using the feature quantity calculating method according to any one of the above embodiments. Therefore, the screening of the pharmaceutical candidate compound can be efficiently performed.

[0028] In the seventh aspect, it is preferable the feature quantities of compounds for which the similarity is determined are calculated according to the same aspect (using the same feature quantity calculating method). The same applies to the various aspects described below.

[0029] According to an eighth aspect of the present invention, there is provided a screening method including: an input step of inputting information related to three-dimensional structures of a plurality of compounds; a first feature quantity calculating step of calculating the feature quantity of each of the plurality of compounds using the feature quantity calculating method according to any one of the above embodiments based on the information; a second feature quantity calculating step of calculating a structural feature quantity indicating a compound structure of each of the plurality of compounds based on the information; and an extracting step of extracting compounds in which a similarity between the feature quantities satisfies a target value and a similarity between the structural feature quantities is less than or equal to a threshold, from the plurality of compounds.

[0030] In order to ensure that a novel compound obtained by search has a skeleton different from the skeleton of the original (existing) compound serving as the starting point of the search, it is effective to impose a condition that "the feature quantities

(according to embodiments of the present invention) of the original compound and the novel compound are the same as each other, but the structural formulae thereof are different from each other" in a case of the search of the compound. The eighth aspect defines one aspect of the screening method based on such an idea, and compounds in which a similarity between the feature quantities satisfies the target value and a similarity between the structural feature quantities is less than or equal to the threshold are extracted from the plurality of compounds using the feature quantity calculating method according to any one of the above embodiments. Therefore, the screening of the pharmaceutical candidate compound can be efficiently performed. Further, it is effective to use the index of the similarity between the structural formulae to indicate that "the structural formulae thereof are different from each other", but many known indices which have been suggested can be used as such an index.

[0031] According to a ninth aspect of the present invention, there is provided a screening method including: an input step of inputting information related to three-dimensional structures of a plurality of compounds; a feature quantity calculating step of calculating the feature quantity of each of the plurality of compounds using the feature quantity calculating method according to any one of the above embodiments based on the information; a predicting step of predicting a binding force using a prediction model that outputs the binding force between each of the plurality of compounds and another compound based on the feature quantity; and an extracting step of extracting compounds, in which the binding force satisfies a target value, from the plurality of compounds. In the screening method according to the ninth aspect, compounds in which the binding force satisfies the target value can be efficiently extracted (screening of the pharmaceutical candidate compound can be efficiently performed) by changing the information serving as the starting point of the search to information related to "the prediction model of the binding force based on the feature quantity of the present invention" and "the target value of the binding force (which can be set to be higher than the current value)" from the information related to only the structural formula of one compound.

[0032] According to a tenth aspect of the present invention, there is provided a screening method including: an input step of inputting information related to three-dimensional structures of a plurality of compounds; a setting step of setting a target value of a binding force between each of the plurality of compounds and another compound; an estimating step of estimating the feature quantity corresponding to the target value using a prediction model that outputs the binding force based on the feature quantity calculated by the feature quantity calculating method according to any

one of the above embodiments; a feature quantity calculating step of calculating the feature quantity of each of the plurality of compounds using the feature quantity calculating method according to any one of the above embodiments based on the information; and an extracting step of extracting compounds in which a similarity between the calculated feature quantity and the estimated feature quantity is greater than or equal to a threshold, from the plurality of compounds based on the calculation results. In the ninth aspect described above, the compound satisfying the "target value of the binding force" is extracted by predicting the binding force using the "prediction model". Meanwhile, in the screening method according to the tenth aspect, the "prediction model" is used in the opposite manner to that of the ninth aspect. Specifically, the feature quantity (the feature quantity according to any one of the above embodiments) is estimated from the "target value of the binding force" using the "prediction model" to search a compound having the estimated feature quantity. Even in a case of using such a method, screening of the pharmaceutical candidate compound can be efficiently performed as in the twelfth aspect.

[0033] Further, in the ninth and tenth aspects, the "prediction model of the binding force" can be generated from, for example, a set of structural formulae of a plurality of binding compounds and experimental values of the binding force thereof.

[0034] According to an eleventh aspect of the present invention, there is provided a screening program which causes a computer to execute the screening method according to any one of the above embodiments. The "computer" in the eleventh aspect can be realized using one or more of various processors such as a CPU. This "computer" may be a computer comprised in a screening device or a screening system. Further, a non-transitory recording medium on which a computer-readable code of the screening program according to the eleventh aspect is recorded can also be exemplified as an aspect of the present invention.

[0035] According to a twelfth aspect of the present invention, there is provided a compound creating method of creating a three-dimensional structure of a target compound that is bound to a target protein from a plurality of compounds, the method including: a storing step of storing a three-dimensional structure of a compound formed of a plurality of atoms and the first feature quantity calculated using the feature quantity calculating method according to the first embodiment in association with each other for each of the plurality of compounds; a feature quantity calculating step of calculating the first feature quantity of a ligand that is a compound whose binding to the target protein has been confirmed using the feature quantity calculating method according to the first embodiment; a generator constructing step of

constructing a generator through machine learning using the three-dimensional structures of the plurality of compounds as teacher data and the first feature quantity as an explanatory variable; and a compound three-dimensional structure generating step of generating the three-dimensional structure of the target compound from the first feature quantity of the ligand using the generator.

[0036] In the screening methods according to the above aspects, a compound that is compatible with a ligand or a target protein is found among a plurality of compounds whose structural formulae have already been determined (written down). Accordingly, after the feature quantity of the compound is calculated, a method of extracting the compound based on the similarity with the feature quantity of the separately calculated ligand or the pocket structure of the target protein, that is, a search method is employed. Therefore, in a case where the correspondence between the structural formula of the compound and the feature quantity thereof is recorded, a structural formula having a high similarity (greater than or equal to the threshold) can be found. Meanwhile, in the twelfth aspect, a structural formula of a compound having a feature quantity similar to the feature quantity (the first feature quantity) of the ligand (accordingly, the drug efficacies are similar) is generated without performing search.

[0037] The generation of the structural formula in a case where the feature quantity has been provided can be performed using a generator constructed through machine learning. Specifically, in the twelfth aspect, a generator is constructed through machine learning (the learning method is not particularly limited) using the three-dimensional structures of the compounds as teacher data and the first feature quantity as an explanatory variable, and a three-dimensional structure of the target compound is generated from the first feature quantity of the ligand using the generator. In the twelfth aspect, since search is not performed, the three-dimensional structure of the compound can be generated even in a case of “no solution was found as the result of screening search”, and thus the three-dimensional structure of the pharmaceutical candidate compound can be efficiently created. Further, the three-dimensional structure to be generated in the twelfth aspect is affected by the features of the compound provided as teacher data. Therefore, by selecting the features of the compound to be provided as teacher data, a compound having a three-dimensional structure with different features can be generated. For example, a compound having a three-dimensional structure that is easily synthesized can be generated by providing a compound that is easily synthesized as teacher data.

[0038] According to a thirteenth aspect of the present invention, there is provided a compound creating method of creating a three-dimensional structure of a target compound that is bound to a target protein from a plurality of compounds, the method including: a storing step of storing a three-dimensional structure of a compound formed of a plurality of atoms and the first feature quantity calculated using the feature quantity calculating method according to the first embodiment in association with each other for each of the plurality of compounds; a feature quantity calculating step of calculating the second feature quantity of the pocket structure of the target protein using the feature quantity calculating method according to the second embodiment; a generator constructing step of constructing a generator through machine learning using the three-dimensional structures of the plurality of compounds as teacher data and the first feature quantity as an explanatory variable; and a compound three-dimensional structure generating step of generating the three-dimensional structure of the target compound from the second feature quantity of the pocket structure using the generator. According to the thirteenth aspect, similar to the twelfth aspect, a structural formula of a compound having a feature quantity similar to the feature quantity (the second feature quantity) of the pocket structure (accordingly, the drug efficacies are similar) is generated without performing search, and thus the three-dimensional structure of the pharmaceutical candidate compound can be efficiently created. Further, similar to the twelfth aspect, by selecting the features of the compound to be provided as teacher data, a compound having a three-dimensional structure with different features can be generated.

[0039] According to a fourteenth aspect of the present invention, there is provided a compound creating method of creating a three-dimensional structure of a target compound that is bound to a target biopolymer other than a protein from a plurality of compounds, the method including: a storing step of storing a three-dimensional structure of a compound formed of a plurality of atoms and a third feature quantity in association with each other for each of the plurality of compounds; a feature quantity calculating step of calculating the third feature quantity of a binding compound that is a compound whose binding to the target biopolymer other than the protein has been confirmed using the feature quantity calculating method according to the third embodiment; a generator constructing step of constructing a generator through machine learning using the three-dimensional structures of the plurality of compounds as teacher data and the third feature quantity as an explanatory variable; and a compound three-dimensional structure generating step of generating the three-dimensional structure of the target compound from the third feature quantity of the binding compound using the generator.

[0040] According to the fourteenth aspect, similar to the twelfth and thirteenth aspects, a structural formula of a compound having a feature quantity similar to the feature quantity (the third feature quantity) of the binding compound (accordingly, the drug efficacies are similar) is generated without performing search, and thus the three-dimensional structure of the pharmaceutical candidate compound can be efficiently created. Further, similar to the twelfth and thirteenth aspects, by selecting the features of the compound to be provided as teacher data, a compound having a three-dimensional structure with different features can be generated.

[0041] As described above, according to the feature quantity calculating method, the feature quantity calculating program, and the feature quantity calculating device of the present invention, it is possible to calculate a feature quantity that accurately shows chemical properties of a target structure. Further, according to the screening method and the screening program of the present invention, screening of a pharmaceutical candidate compound can be efficiently performed. Further, according to the compound creating method of the present invention, the three-dimensional structure of a pharmaceutical candidate compound can be efficiently created.

BRIEF DESCRIPTION OF THE DRAWINGS

[0042] Fig. 1 is a block diagram showing a configuration of a screening device according to a first embodiment.

Fig. 2 is a block diagram showing a configuration of a processing unit.

Fig. 3 is a diagram showing information stored in a storage unit.

Fig. 4 is a diagram showing a state in which structure information of a compound and a feature quantity thereof are stored in association with each other.

Fig. 5 is a flowchart showing a procedure for calculating a three-dimensional AAM