type of interaction, which is a feature of this technique, needs to be set in advance by a person, and there is uncertainty in the classification of the type of interaction.

[0011] In addition, in the SBDD method described in WO2006/099178A, there is no uncertainty in the classification of the type of interaction in the pharmacophore method, but the calculation of the bonding energy in the MD simulation takes time because information on a target having a large molecular weight is used, and a chemical space of a compound as a target for search and design is limited. In addition, in the AAM method described in JP6826672B, there is no restriction on the chemical space due to an increase in calculation time caused by handling a target in the SBDD method, but there is a problem that a molecular weight hardly changes upon the skeleton conversion.

[0012] As described above, a technique for systematically executing reduction of a molecular weight of an antibody or the like does not currently exist, and there is a demand for development of such a method.

[0013] Embodiments of the present invention provide a compound searching device, a compound searching method, and a compound searching program that can efficiently search for a low-molecular-weight compound having activity and membrane permeability.

[0014] According to a first aspect of the present invention, there is provided a compound searching device comprising a processor, in which the processor is configured to: acquire information indicating a three-dimensional structure of a search target compound; acquire information indicating a conformation of a ligand in a complex structure of the ligand and a target protein; calculate a first feature amount that is a feature amount obtained by quantifying, in a three-dimensional space, a degree of accumulation of one or more amino acids as a probe for the ligand in a periphery of the ligand with respect to the conformation of the ligand; specify, as a main site, a site of a part of the ligand, of which a proportion in a bonding force of the ligand to the target protein is equal to or higher than a first threshold value; calculate a partial feature amount that is a part derived from the main site, in the first feature amount; calculate a second feature amount that is a feature amount obtained by quantifying, in a three-dimensional space, a degree of accumulation of one or more amino acids as a probe for the search target compound in a periphery of the search target compound; calculate a similarity between the second feature amount for the search target compound and the partial feature amount of the ligand; and output the similarity for the search target compound and information indicating a molecular weight and/or activity of the search target compound in association with each other to a recording device and/or a display device.

[0015] According to the above aspect, the “partial feature amount” is a feature amount derived from the main site in a feature amount derived from the entire ligand, and is likely to have a smaller value than the first feature amount. Then, in the above aspect, the similarity between the “partial feature amount” and the second feature amount derived from the entire search target compound is calculated, and a compound having a high similarity (equal to or higher than a second threshold value) is calculated. Since the compound having a high similarity has a feature amount (second feature amount) having a value smaller than (or having a high possibility of being smaller than) the first feature amount, the compound having a high similarity is likely to have a smaller molecular weight than the original ligand. In the above aspect, the similarity of the search target compound and the information indicating the molecular weight and/or the activity of the search target compound are output in association with each other to the recording device and/or the display device.

[0016] As described above, since the reduction in molecular weight is one of the measures for imparting activity and membrane permeability, according to the above aspect, it is possible to efficiently search for a low-molecular-weight compound having activity and membrane permeability. In the above aspect, the processor may extract a compound satisfying a desired condition from the search target compound with reference to the output information (the similarity of the search target compound and the information indicating the molecular weight and/or the activity of the search target compound).

[0017] In the above aspect and each of the following embodiments, the first feature amount for the ligand may be referred to as an “overall feature amount” of the ligand, and the second feature amount for the search target compound may be referred to as an “overall feature amount” of the search target compound.

[0018] In the above aspect and each of the following embodiments, the processor may use a compound that is acquired in advance and recorded on a recording medium as the “search target compound”, or may search for a compound recorded in a library each time the processing is performed. In addition, the processor may specifically designate the search target compound (according to an operation of a user or without depending on the operation of the user) in the processing. Such a recording medium and library may be included in the compound searching device according to the above aspect, or may be recording media and libraries outside the device.

[0019] In addition, in the above aspect and each of the following embodiments, the “information indicating a three-dimensional structure” may be the three-dimensional structure

itself or may be information from which the three-dimensional structure can be indirectly obtained. For example, in a case where a structural formula is acquired as the “information indicating a three-dimensional structure”, the three-dimensional structure of the compound can be obtained from the structural formula by various methods.

[0020] According to an embodiment, in the above compound searching device , the processor is configured to, in specifying the main site, acquire any one of a result of bonding energy calculation using a complex structure of the ligand and the target protein, information indicating a structural activity correlation between the ligand and a ligand analogous substance, or the first feature amount for a local stable conformation of the ligand from a non-transitory and tangible recording medium, as the information indicating the bonding force, and specify the main site based on the acquired information. This embodiment defines one embodiment of a technique for specifying the main site.

[0021] In the above embodiment, examples of the “non-transitory and tangible recording medium” include various magneto-optical recording devices and semiconductor memories. The “non-transitory and tangible recording medium” does not include a non-tangible recording medium such as a carrier wave signal itself and a propagation signal itself.

[0022] According to an embodiment, in the above compound searching device, the processor is configured to, in specifying the main site, select one or more amino acid residues in which a ratio of a total value of a breakdown of a bonding force due to each amino acid residue of the ligand to the bonding force of the ligand is equal to or higher than the first threshold value, and specify a site of the one or more amino acid residues in the ligand as the main site. This embodiment specifically defines one embodiment of a technique for specifying the main site.

[0023] According to an embodiment, in the above compound searching device, the processor is configured to, in specifying the main site, select the one or more amino acid residues in consideration of a distance between the amino acid residues. This embodiment more specifically defines the technique for specifying the main site in the above embodiment.

[0024] According to an embodiment, in the above compound searching device, the processor is configured to, in specifying the main site, specify a site of the ligand present in a designated region as the main site. In this embodiment, the processor may designate the region according to the operation of the user, or may designate the region without depending on the operation of the user.

[0025] According to an embodiment, in the above compound searching device, the processor is configured to, in specifying the main site, designate the region based on a distribution of the

first feature amount in a periphery of the ligand. In this embodiment, for example, the processor can designate a region including a portion in which the degree of accumulation is maximized.

[0026] According to an embodiment, in the above compound searching device, the processor is configured to, in specifying the main site, determine at least one of the number, position, size, or shape of the region according to an operation of a user via an input device. According to this embodiment, the user can designate at least one of the number, position, size, or shape of the region.

[0027] According to an embodiment, in the above compound searching device, the processor is configured to, in specifying the main site, display, on the display device, information indicating a three-dimensional structure of the ligand, a distribution of the first feature amount in a periphery of the ligand, and the number, position, size, and shape of the region in a superimposed manner. According to this embodiment, the user can easily ascertain a relationship between the degree of accumulation and the region by this display.

[0028] According to an embodiment, in the above compound searching device, the processor is configured to display information indicating a ratio of the partial feature amount to the first feature amount on the display device. According to this embodiment, the user can easily ascertain the ratio of the partial feature amount to the first feature amount.

[0029] According to an embodiment, in the above compound searching device, the processor is configured to, in specifying the main site, display, on the display device, the site of the one or more amino acid residues specified as the main site in a three-dimensional structure of the ligand in an identifiable manner. According to this embodiment, the user can easily ascertain the site of the amino acid residue specified as the main site.

[0030] According to an embodiment, in the above compound searching device, the processor is configured to, in specifying the main site, calculate a breakdown of bonding energy between the ligand and the target protein for each amino acid residue constituting the ligand, and display the calculated breakdown on the display device for each amino acid residue. According to this embodiment, the user can easily ascertain the bonding energy by each residue.

[0031] According to an embodiment, in the above compound searching device, the processor is configured to: calculate a plurality of partial feature amounts by translating and/or rotating the partial feature amount of the ligand in the calculation of the partial feature amount; and calculate the similarity based on a cosine similarity or a Euclidean distance between the

plurality of partial feature amounts and the second feature amount for the search target compound in the calculation of the similarity. The embodiment specifically defines one technique for calculating the similarity.

[0032] According to an embodiment, in above the compound searching device, the processor is configured to, in the outputting, extract a compound of which the similarity is equal to or higher than a second threshold value from the search target compounds, and output information indicating the extracted compound to the recording device and/or the display device. In this embodiment, the processor may extract a predetermined number of compounds in descending order of the similarity, or may extract all the compounds having a similarity equal to or higher than the second threshold value. The processor may determine the number of compounds to be extracted according to the operation of the user.

[0033] According to a second aspect of the present invention, there is provided a compound searching method executed by a compound searching device including a processor, the method comprising: via the processor, acquiring information indicating a three-dimensional structure of a search target compound; acquiring information indicating a conformation of a ligand in a complex structure of the ligand and a target protein; calculating a first feature amount that is a feature amount obtained by quantifying, in a three-dimensional space, a degree of accumulation of one or more amino acids as a probe for the ligand in a periphery of the ligand with respect to the conformation of the ligand; specifying, as a main site, a site of a part of the ligand, of which a proportion in a bonding force of the ligand to the target protein is equal to or higher than a first threshold value; calculating a partial feature amount that is a part derived from the main site, in the first feature amount; calculating a second feature amount that is a feature amount obtained by quantifying, in a three-dimensional space, a degree of accumulation of one or more amino acids as a probe for the search target compound in a periphery of the search target compound; calculating a similarity between the second feature amount for the search target compound and the partial feature amount of the ligand; and outputting the similarity for the search target compound and information indicating a molecular weight and/or activity of the search target compound in association with each other to a recording device and/or a display device.

[0034] According to the above aspect, similarly to the first aspect, it is possible to efficiently search for a low-molecular-weight compound having activity and membrane permeability. The compound searching method according to the above aspect may have the same configurations as those of the second to thirteenth aspects.

[0035] Disclosed herein is a compound searching program for causing a compound searching device including a processor to execute a compound searching method, the compound searching method comprising: via the processor, acquiring information indicating a three-dimensional structure of a search target compound; acquiring information indicating a conformation of a ligand in a complex structure of the ligand and a target protein; calculating a first feature amount that is a feature amount obtained by quantifying, in a three-dimensional space, a degree of accumulation of one or more amino acids as a probe for the ligand in a periphery of the ligand with respect to the conformation of the ligand; specifying, as a main site, a site of a part of the ligand, of which a proportion in a bonding force of the ligand to the target protein is equal to or higher than a first threshold value; calculating a partial feature amount that is a part derived from the main site, in the first feature amount; calculating a second feature amount that is a feature amount obtained by quantifying, in a three-dimensional space, a degree of accumulation of one or more amino acids as a probe for the search target compound in a periphery of the search target compound; calculating a similarity between the second feature amount for the search target compound and the partial feature amount of the ligand; and outputting the similarity for the search target compound and information indicating a molecular weight and/or activity of the search target compound in association with each other to a recording device and/or a display device.

[0036] According to the above disclosure, similarly to the first and second aspects, it is possible to efficiently search for a low-molecular-weight compound having activity and membrane permeability. The compound searching program according to the above disclosure may have the same configuration as those of the second to thirteenth aspects.

[0037] In addition, according to a third aspect, there is provided a non-transitory and tangible recording medium (for example, various magneto-optical recording devices and semiconductor memories) on which a computer-readable code of a program for causing a processor of a computer to execute the compound searching method of above aspects is recorded. The “non-transitory and tangible recording medium” described above does not include a non-tangible recording medium such as the carrier wave signal itself and the propagation signal itself.

[0038] As described above, according to the compound searching device, the compound searching method, and the compound searching program of the above aspects of the present invention, it is possible to efficiently search for a low-molecular-weight compound having activity and membrane permeability.

BRIEF DESCRIPTION OF THE DRAWINGS

[0039] Fig. 1 is a diagram showing a configuration of a compound searching device.

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

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

Fig. 4 is a flowchart showing processing of a compound searching method.

Fig. 5 is a flowchart showing a method of calculating a three-dimensional AAM descriptor.

Figs. 6A and 6B are diagrams showing a state in which a structural formula is three-dimensionalized.

Figs. 7A and 7B are diagrams showing an example of the three-dimensional AAM descriptor.

Fig. 8 is another diagram showing an example of the three-dimensional AAM descriptor.

Figs. 9A, 9B, and 9C are diagrams showing a state in which a main site is specified.

Fig. 10 is a diagram showing an example of display of a contribution rate of an interaction due to a selected residue.

Figs. 11A, 11B, and 11C are diagrams showing a state in which a partial feature amount is calculated.

Figs. 12A, 12B, and 12C are diagrams showing a state in which an overall feature amount of a search target compound is calculated.

Figs. 13A and 13B are diagrams showing a state in which a similarity between feature amounts is calculated.

Fig. 14 is a diagram showing names and molecular weights of the search target compounds.

Fig. 15 is a diagram showing a relationship between a similarity of a feature amount and an activity.

Figs. 16A and 16B are diagrams showing a similarity between an overall feature amount of a ligand and a partial feature amount and an overall feature amount of the search

three-dimensional space, the degree of accumulation of one or more kinds of amino acids "a" in the periphery of the three-dimensional structure of the compound. The feature amount calculation unit 118 can store the calculated three-dimensional AAM descriptor in the storage unit 200 in association with the structural information (information indicating the three-dimensional structure) of the compound (see Fig. 3). In addition, "r" in $\Delta G_{a\mu}(r)$, $g_{a\mu}(r)$, and $g_a(r)$ is a three-dimensional position vector ($r = (x, y, z)$).

[0066] Figs. 7A and 7B are an example of a three-dimensional AAM descriptor of the compound shown in Figs. 6A and 6B. Fig. 7A shows a three-dimensional AAM descriptor for alanine, and Fig. 7B shows a three-dimensional AAM descriptor for valine. In Figs. 7A and 7B, a dark region is a region where the degree of accumulation of amino acids (existence probability) is high. Fig. 8 is a table showing another example of the three-dimensional AAM descriptor for the compound shown in Figs. 7A and 7B, and shows the three-dimensional AAM descriptors in a direction 1, a direction 2, and a direction 3 (first, second, and third rows of the table, respectively) which are different from one another. A left column of the table shows the three-dimensional AAM descriptor (level surface with respect to a threshold value), and a right column of the table shows the three-dimensional AAM descriptor (level surface with respect to the threshold value) and the three-dimensional structure of the compound.

[0067] [Specification of Main Site]

The main site specifying unit 116 (processor) specifies a site of a part of the ligand, of which a proportion in a bonding force of the ligand to the target protein is equal to or higher than a first threshold value, as a main site (step S120 in Fig. 4: main site specifying processing, main site specifying step).

[0068] The main site specifying unit 116 can specify the main site by, for example, any one of the following methods (1) to (3). The main site specifying unit 116 may specify the main site without using the "three-dimensional AAM descriptor" as in the method (1) and the method (2), or may specify the main site using the "three-dimensional AAM descriptor" as in the method (3).

(1) Using results of bonding energy calculation using a complex structure of a ligand and a target protein (using FMO, MO, MD, MM, or the like) (FMO method: Fragment Molecular Orbital method, MO method: Molecular Orbital method).

(2) Using a calculation result of a structural activity correlation between a ligand and a ligand analogous substance.

feature amount at any angle are calculated, all of cos similarities between the partial feature amounts are calculated, and a maximum value among the cos similarities is calculated as the "similarity".

(4) A three-dimensional AAM descriptor (overall feature amount (second feature amount) of the search target compound) is calculated for a plurality of local stable conformations obtained by executing search calculation (MD, MM, AlphaFold2, or the like) of the local stable conformation of the molecule, a plurality of partial feature amounts created by translating the partial feature amount of the ligand at any distance and rotating the partial feature amount at any angle are calculated, all of Euclid distances between the partial feature amounts are calculated, and a minimum value among the Euclid distances is calculated as the "similarity".

[0076] The similarity calculation unit 122 may determine which method is used to calculate the similarity by the operation of the user via the operation unit 400, or may determine the method without depending on the operation of the user.

[0077] [Output of Relationship between Similarity and Molecular weight/Activity]

The input/output control unit 126 (processor) outputs the similarity for the search target compound and information indicating the molecular weight and/or the activity of the search target compound in association with each other to the recording device (storage unit 200 and the like) and/or the display device (monitor 310 and the like) (step S170: output step, output processing). The processor 110 may acquire information already calculated for the molecular weight and the activity, or may calculate the molecular weight and the activity from the structural information of the search target compound.

[0078] As described above, according to the compound searching device 10 (compound searching device), the compound searching method, and the compound searching program according to the first embodiment, it is possible to provide information for efficiently extracting a compound reduced in molecular weight. As described above, since the reduction in molecular weight is one of the measures for imparting activity and membrane permeability, according to the first embodiment, it is possible to efficiently search for a low-molecular-weight compound having activity and membrane permeability. Specific output examples will be described later in Examples (see Figs. 15 and 16, and the like).

[0079] [Extraction of Compound]

In addition, the compound extraction unit 124 (processor) may extract a compound satisfying a desired condition with reference to the calculated information (similarity,

CLAIMS

1. A compound searching device comprising a processor,

wherein the processor is configured to:

acquire information indicating a three-dimensional structure of a search target compound;

acquire information indicating a conformation of a ligand in a complex structure of the ligand and a target protein;

calculate a first feature amount that is a feature amount obtained by quantifying, in a three-dimensional space, a degree of accumulation of one or more amino acids as a probe for the ligand in a periphery of the ligand with respect to the conformation of the ligand;

specify, as a main site, a site of a part of the ligand, of which a proportion in a bonding force of the ligand to the target protein is equal to or higher than a first threshold value;

calculate a partial feature amount that is a part derived from the main site, in the first feature amount;

calculate a second feature amount that is a feature amount obtained by quantifying, in a three-dimensional space, a degree of accumulation of one or more amino acids as a probe for the search target compound in a periphery of the search target compound;

calculate a similarity between the second feature amount for the search target compound and the partial feature amount of the ligand; and

output the similarity for the search target compound and information indicating a molecular weight and/or activity of the search target compound in association with each other to a recording device and/or a display device.

2. The compound searching device according to claim 1,

wherein the processor is configured to, in specifying the main site,

acquire any one of a result of bonding energy calculation using a complex structure of the ligand and the target protein, information indicating a structural activity correlation between the ligand and a ligand analogous substance, or the first feature amount for a local stable conformation of the ligand from a non-transitory and tangible recording medium, as the information indicating the bonding force, and

specify the main site based on the acquired information.

3. The compound searching device according to claim 1 or 2,

wherein the processor is configured to, in specifying the main site,

select one or more amino acid residues in which a ratio of a total value of a breakdown of a bonding force due to each amino acid residue of the ligand to the bonding force of the ligand is equal to or higher than the first threshold value, and specify a site of the one or more amino acid residues in the ligand as the main site.

4. The compound searching device according to claim 3,

wherein the processor is configured to, in specifying the main site,

select the one or more amino acid residues in consideration of a distance between the amino acid residues.

5. The compound searching device according to any one of claims 1 to 4,

wherein the processor is configured to, in specifying the main site,

specify a site of the ligand present in a designated region as the main site.

6. The compound searching device according to claim 5, wherein the processor is configured to, in specifying the main site,

designate the region based on a distribution of the first feature amount in a periphery of the ligand.

7. The compound searching device according to claim 5,

wherein the processor is configured to, in specifying the main site,

determine at least one of the number, position, size, or shape of the region according to an operation of a user via an input device.

8. The compound searching device according to any one of claims 5 to 7,

wherein the processor is configured to, in specifying the main site,

display, on the display device, information indicating a three-dimensional structure of the ligand, a distribution of the first feature amount in a periphery of the ligand, and the number, position, size, and shape of the region in a superimposed manner.

9. The compound searching device according to any one of claims 1 to 8,

wherein the processor is configured to display information indicating a ratio of the

partial feature amount to the first feature amount on the display device.

10. The compound searching device according to any one of claims 3 to 9,

wherein the processor is configured to, in specifying the main site,

display, on the display device, the site of the one or more amino acid residues specified as the main site in a three-dimensional structure of the ligand in an identifiable manner.

11. The compound searching device according to any one of claims 1 to 10,

wherein the processor is configured to, in specifying the main site,

calculate a breakdown of bonding energy between the ligand and the target protein for each amino acid residue constituting the ligand, and display the calculated breakdown on the display device for each amino acid residue.

12. The compound searching device according to any one of claims 1 to 11,

wherein the processor is configured to:

calculate a plurality of partial feature amounts by translating and/or rotating the partial feature amount of the ligand in the calculation of the partial feature amount; and

calculate the similarity based on a cosine similarity or a Euclidean distance between the plurality of partial feature amounts and the second feature amount for the search target compound in the calculation of the similarity.

13. The compound searching device according to any one of claims 1 to 12,

wherein the processor is configured to, in the outputting,

extract a compound of which the similarity is equal to or higher than a second threshold value from the search target compounds, and output information indicating the extracted compound to the recording device and/or the display device.

14. A compound searching method executed by a compound searching device including a processor, the method comprising:

via the processor,

acquiring information indicating a three-dimensional structure of a search target compound;

acquiring information indicating a conformation of a ligand in a complex structure of

the ligand and a target protein;

calculating a first feature amount that is a feature amount obtained by quantifying, in a three-dimensional space, a degree of accumulation of one or more amino acids as a probe for the ligand in a periphery of the ligand with respect to the conformation of the ligand;

specifying, as a main site, a site of a part of the ligand, of which a proportion in a bonding force of the ligand to the target protein is equal to or higher than a first threshold value;

calculating a partial feature amount that is a part derived from the main site, in the first feature amount;

calculating a second feature amount that is a feature amount obtained by quantifying, in a three-dimensional space, a degree of accumulation of one or more amino acids as a probe for the search target compound in a periphery of the search target compound;

calculating a similarity between the second feature amount for the search target compound and the partial feature amount of the ligand; and

outputting the similarity for the search target compound and information indicating a molecular weight and/or activity of the search target compound in association with each other to a recording device and/or a display device.

15. A non-transitory computer-readable recording medium on which a program for causing a processor of a computer to execute the compound searching method according to claim 14 is recorded.