(12) PATENT **(11)** Application No. **AU 199919722 B2**
(19) AUSTRALIAN PATENT OFFICE **(10)** Patent No. **771430**

(54) Title
Method for providing, identifying and describing molecules capable of having a required activity, in particular in pharmacology and molecules resulting from said method

(51)[7] International Patent Classification(s)
G06F 017/50 **G06F 019/00**

(21) Application No: **199919722** (22) Application Date: **1998.12.29**

(87) WIPO No: **WO99/35598**

(30) Priority Data

(31) Number	(32) Date	(33) Country
97/16706	**1997.12.30**	**FR**

(43) Publication Date : **1999.07.26**
(43) Publication Journal Date : **1999.09.30**
(44) Accepted Journal Date : **2004.03.25**

(71) Applicant(s)
Synt:em (S.A.)

(72) Inventor(s)
Gerard Grassy; Michel Kaczorek; Roger Lahana; Abdelaziz Yasri

(74) Agent/Attorney
Griffith Hack,GPO Box 1285K,MELBOURNE VIC 3001

(56) Related Art
WO 1997/027559

19722/99

PCT ORGANISATION MONDIALE DE LA PROPRIETE INTELLECTUELLE
Bureau international

DEMANDE INTERNATIONALE PUBLIEE EN VERTU DU TRAITE DE COOPERATION EN MATIERE DE BREVETS (PCT)

(51) Classification internationale des brevets 6 : G06F 17/50 , 19/00 , 159/00

A1

(11) Numéro de publication internationale: **WO 99/35598**

(43) Date de publication internationale: 15 juillet 1999 (15.07.99)

(21) Numéro de la demande internationale: PCT/FR98/02909

(22) Date de dépôt international: 29 décembre 1998 (29.12.98)

(30) Données relatives à la priorité:
97/16706 30 décembre 1997 (30.12.97) FR

(71) Déposant *(pour tous les Etats désignés sauf US)*: SYNT:EM (S.A.) [FR/FR]; Parc Scientifique, Georges Besse, F-30000 Nîmes (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants *(US seulement)*: GRASSY, Gérard [FR/FR]; 32, rue du Pradas, F-34470 Pérols (FR). KACZOREK, Michel [FR/FR]; Synt:Em (S.A.), 145, allée Charles Babbage, F-30900 Nîmes (FR). LAHANA, Roger [FR/FR]; Synt:Em (S.A.), 145, allée Charles Babbage, F-30900 Nîmes (FR). YASRI, Abdelaziz [FR/FR]; Synt:Em (S.A.), 145, allée Charles Babbage, F-30900 Nîmes (FR).

(74) Mandataire: BREESE, Pierre; Breese-Majerowicz, 3, avenue de l'Opéra, F-75001 Paris (FR).

(81) Etats désignés: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, brevet ARIPO (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), brevet eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Publiée

Avec rapport de recherche internationale.
Avant l'expiration du délai prévu pour la modification des revendications, sera republiée si des modifications sont reçues.

(54) Title: METHOD FOR PROVIDING, IDENTIFYING AND DESCRIBING MOLECULES CAPABLE OF HAVING A REQUIRED ACTIVITY, IN PARTICULAR IN PHARMACOLOGY AND MOLECULES RESULTING FROM SAID METHOD

(54) Titre: PROCEDE POUR PREVOIR, IDENTIFIER ET DECRIRE DES MOLECULES SUSCEPTIBLES DE PRESENTER UN COMPORTEMENT RECHERCHE, NOTAMMENT DANS LE DOMAINE DE LA PHARMACIE ET MOLECULES OBTENUES PAR CE PROCEDE

(57) Abstract

The invention concerns a computer-assisted method for providing, identifying and describing molecules having a required activity, in particular in pharmacology, comprising a modelling step, a step for constituting a combinatorial bank and a step for selecting potentially interesting molecules, including a step of dynamic filtering of the candidate molecules representative of the conformation-varying stresses which the molecules must observe to have said activity.

(57) Abrégé

La présente invention concerne un procédé assisté par ordinateur pour prévoir, identifier et décrire des molécules ayant un comportement recherché, notamment dans le domaine de la pharmacie, mettant en oeuvre une étape de modélisation moléculaire, une étape de constitution d'une banque combinatoire et une étape de sélection des molécules potentiellement intéressantes, comportant une étape de filtration des molécules candidates par un filtre dynamique représentatif des contraintes de variations conformationnelles que doivent respecter les molécules pour présenter ladite activité.

METHOD FOR PROVIDING, IDENTIFYING AND DESCRIBING MOLECULES CAPABLE OF HAVING A REQUIRED ACTIVITY, IN PARTICULAR IN PHARMACOLOGY, AND MOLECULES RESULTING FROM SAID METHOD

This invention relates to molecular modeling.

Its purpose is a new process using given number of initial molecules, particularly but not only by virtual combinatorial chemistry, in order to provide, identify and describe molecules capable of having a required activity, in particular in pharmacology.

It is known that the search for new active molecules, particularly in pharmacology, requires synthesis of a very large number of molecules that have to be tested *in vitro* or *in vivo*. At most, only a very small number of these molecules proves to be active.

Molecular modeling with computerised data bases was used in order to rationalise the search for new active molecules. Results obtained with this type of technique in the past have not always been satisfactory, particularly due to an inadequate definition of parameters and activity criteria.

It was known in the prior art the International patent application WO/9727559 which discloses a method to select molecules by measuring the distance between two descriptors.

It was also known the document YASRI ET AL: <<rational choice of molecular dynamics simulation parameters through the use of the three-dimensional aurocorrelation method: application to calmodulin flexibility study>> PROTEIN ENGINEERING., vol. 9, no. 11, November, 1996, pages 959-976 XP002078077 England GB.

The present invention provides, therefore a computer-assisted process for providing, identifying and describing molecules with a required activity, involving:

1) providing a learning base composed firstly of active molecular structures known to have the required activity and secondly of inactive molecular structures

known not to have this activity or only to express this activity very weakly, using appropriate descriptors;

2) generating static filters representative of the intervals of structural, physicochemical and molecular values that must be satisfied by the molecules if they are to be active, starting from the learning base;

3) screening a combinatorial bank of molecules on the basis of said static filters; and

4) submitting the most promising molecules thus selected to dynamic filtering representative of conformational variation constraints which the molecules must observe in order to be active.

The term "activity" means an "activity" in the biological or pharmacological meaning, when the molecules concern pharmaceutical applications, or a "property" in the physicochemical meaning when the molecules concern non-pharmaceutical applications, for example for materials such as polymers.

The invention also provides a process for providing, identifying and describing molecules with a required activity, involving:

1) creating a learning base starting from a family of known similar molecular structures composed firstly of active molecular structures known to have the required activity and secondly of inactive molecular structures known to not have this activity or to not have this activity at all or only very little, using appropriate descriptors;

2) generating a combinatorial explosion of molecules starting from the learning base;

3) screening the molecules thus generated on the enrichment base, in terms of molecular diversity provided by each molecule regarding the chosen descriptors;

4) submitting the selected molecules to static filters representative of the structural, physicochemical and molecular variation constraints that must be satisfied

by the molecules if they are to be active, and possibly to synthesize and test the selected molecules;

5) submitting the most promising molecules thus selected to dynamic filtering representative of confirmation variation constraints which the molecules must observe in order to be active; and

6) synthesizing and testing the molecules thus selected.

The invention also relates to a molecule that was not previously known to have a required activity characterized in that it occupies a conformational space identical to or very similar to the conformational space of at least one reference molecule, the reference molecule being previously known to have the required activity. The "very similar" concept is evaluated as a function of all conformational spaces corresponding to the tested molecules. For example, the 10% of all conformational spaces closest to the conformational space of reference molecules will be considered as being "very similar".

When the process is based on several reference molecules, conformational spaces of the reference molecules will be considered as being "very similar".

A learning base is made using various descriptors that may relate to any type of quantitative and/or semi-quantitative properties. For example, as is known to those skilled in the art, a set of numeric values called "topological descriptors" may be deduced starting from a graphic representation of a molecule, in other words a two-dimensional (2D) representation of a chemical compound. Descriptors reflecting some physicochemical properties, for example such as the lipophile character or lipophilia may also be used, which is expressed as logP, where P is the partition coefficient of the compound considered between water and n-octanol, or molar refractivity. Numeric descriptors representing molecular forms can also be used.

The techniques normally used in Quantitative Structure Activity Relationships (QSAR) are based on the assumption that if a molecule has a given biological activity, all information necessary to characterize it is strongly based on its structure, in other words its atoms, links and forms. Unlike the conventional linear QSAR paradigm in which the biological activity can be expressed in the form of a linear combination of relevant descriptors, the Applicant uses a variable non-linear mapping or information representation paradigm in which the activity is a non-linear function of structural, topological and molecular descriptors.

According to the invention, a set of descriptors may be analysed for active compounds, as opposed to inactive compounds, so that filters can be defined to differentiate between the two classes of molecules [G. Grassy et. al., J. Mol. Graphics, vol. 13, page 356 (1995)]

If the learning sample corresponds to the *in vivo* tests, the set of filters describes *in vivo* activity conditions. In the case of *in vitro* tests, the set of filters describes *in vitro* activity conditions.

Preferably a filter is defined by the variation range for a given descriptor for all known active compounds when it is compared with the variation range for the same descriptor for all known inactive compounds.

If the "active" range completely covers the "inactive" range, the filter is not useful, which is equivalent to saying that the variation of this particular descriptor is unrelated to the biological activity.

If the "active" range only partially covers the "inactive" range, any molecule with a value for this descriptor within an area of the "active" range and which does not cover the "inactive" area has a strong probability that it will also be active.

Predictability in this type of approach is improved by the use of several filters, preferably between 10 and 30, each representing a different manner of quantifying the structural and physicochemical properties of a molecule.

Once these filters have been defined, the can be used to screen the virtual compounds generated during a combinatorial explosion, in other words all compounds obtained for example by connecting a list of substitutes on predetermined parts of a structural core called a

synthon. The most promising compounds, namely compounds that satisfy all constraints defined by the filters, are then synthesised by chemical synthesis or genetic engineering, or by any other means, and tested in biological tests.

Inactive molecules that were predicted as being active contain unexplored characteristics which are controlling for the activity. This type of property is used for the definition of additional filters and after a few

repetitions, the virtual screening process may be used for a precise identification of compounds with the required properties.

In practice, a "learning base" is created starting from initial structures, for which the activity may be measured *in vitro* and/or *in vivo*, by determining the various physicochemical, structural and molecular parameters that can be used to describe active molecules and inactive molecules.

These parameters are represented in the form of variation ranges or "constraints" as a function of the activity class. A set of constraints defines a filter.

The quality of the chosen descriptors is checked in terms of variability, together with the choice of descriptors to be used and the intercorrelation between the selected descriptors.

This can be done using software designed by R. Lahana, namely the ANODA (ANalysis Of DAta) software that combines a chemicals library with a "business card" that has simple basic descriptors and analysis descriptors as its main components. Different techniques are used for selection of descriptors such as analyses of the variability, intercorrelation, representativeness (regression, neuron networks, genetic algorithms).

A combinatorial explosion is generated starting from the learning base and making use of all useful variation conditions for each variable position. In doing this, data bases of substitutes defined by their structure and the resulting descriptors can be used.

The combinatorial explosion can be generated using the LEGION software written by the TRIPOS company or another software designed by R. Lahana, the COMBEX software that creates this type of explosion starting from a scaffold or a consensus sequence, and a list of constraints on each substitution point. An SQL (Standard Query Language) type command language can be used to combine all types of substitute selection conditions at will.

For each generated combination, the enrichment in terms of molecular diversity provided by the created molecule with regard to the selected descriptors, can be verified. The

molecule will be kept if this enrichment is sufficient. Otherwise it will be rejected.

Enrichment in terms of molecular diversity may be verified using software designed by R. Lahana, the DIVERSER software that quantitatively evaluates the molecular diversity of any chemicals library, even if the molecules in it are highly flexible and if this chemicals library is entirely virtual. This software can be used to compare banks in terms of diversity, to characterize diversity "holes" in a bank, and to rationally design minimum banks with a maximum diversity. For example, selection techniques may include a hierarchical classification (clusters), partitions, uniform design, random draws, etc.

"Static" filters, in other words filters that do not depend on the conformational variability of molecules, describe intervals of structural, physicochemical and molecular values that must be satisfied by the molecules if they are to be active. By using these static filters, the molecules with the best chances of being active are deduced. These molecules may be synthesized if necessary and tested for the required activity.

Software designed by G. Grassy, the VARIMAP software, can be used in this step to create static filters starting from descriptors obtained by TSAR, which is a software marketed by the Oxford Molecular Group, Oxford (United Kingdom), which was originally written by G. Grassy and R. Lahana.

The most promising candidates selected using static filters are then passed through a dynamic filter, in other words a filter representative of conformational variation constraints that must be respected by the generated molecules if they are to have the best chances of being active, by comparison with molecules in the learning base.

This dynamic filtering may be done using software recently designed by G. Grassy, namely the MULTIDYN software, which uses molecular dynamic paths to characterize conformational spaces of arbitrary molecules. Bioactive conformations of the molecules concerned are thus identified.

The selected molecules are then synthesized and tested.

If the biological activity found does not satisfy the required criteria, or if it only partially satisfies them, the enrichment verification steps in terms of molecular diversity, static filtering, dynamic filtering, synthesis and tests are repeated.

If it is found that the molecules satisfy the defined biological criteria, the purpose has been achieved.

Therefore, the process according to the invention can be used to define new active molecules using an entirely rational process that makes no assumptions at any time about their possible action mechanism or their receiver, if any.

The process according to the invention was successfully used for pharmaco-modulation of molecules in order to improve the performances of these molecules *in vitro* and especially *in vivo*.

The detailed example embodiment described below helps to illustrate and better explain the invention.

EXAMPLE

Obtaining new immuno-modulating compounds

It was recently shown that peptide 2702.75-84 (a peptide derived from HLA-B2702, amino acids 75 to 84) prolongs survival of mice after an allo-heart graft (*Transplantation*, vol. 59, page 455 (1995).

Administration of peptide 2702.75-84 at a dose of 80 mg/kg/day for 10 days after transplant of B6 hearts into CBA receivers prolongs survival following the allo-heart graft up to 11.4 2.6 (n = 8) days compared with 8.2 1.2 days in untreated control animals ($p < 0.01$). No effect on survival following the graft was observed after therapy with peptide 2702.75-84 at lower doses.

The *in vivo* activity of several peptides (n = 19) derived from peptide 2702.75-84 and other MHC/HLA sequences was evaluated in a similar manner. The tested peptides comprised D and L amino acid type peptides. Some peptides differed from peptide 2702.75-84 on up to 6 amino acid positions, and others had an inverse sequence.

All these peptides, for which the sequences are shown in table I below, were synthesized using F-moc/tBu chemistry and were then tested.

Responses of mice that had received the allo-heart graft, to doses, were then studied using peptides in acetate form, purified to more than 90% pure by high performance liquid chromatography (HPLC).

The abdominal heterotropic cardiac transplant was made as described previously by J. Thorac. Cardiovasc. Surg. Vol. 7, page 225 (1969).

CBA mice receiving C57B1/6 hearts were treated daily with different peptide doses after the organ transplant. Peptides were dissolved in DMSO and diluted in PBS (final DMSO concentration 10%) before intraperitoneal administration. The animals were treated starting from the day of the transplant until the 9^{th} day. Survival after the graft was monitored daily by direct palpation and rejection was defined as being the end of palpable cardiac contractility. The statistical significance of prolonged survival following the allo-heart graft was calculated using the Mann-Whitney test.

The therapy of mice receiving allo-heart grafts with some of these peptides (n = 9, shown in bold in table I below) enabled a significant prolongation to survival following the graft, whereas the other peptides had no significant effect.

The structures of the peptides used and the results obtained are shown in table I below.

TABLE I

Peptide sequence	HLA/MHC	MSTSD
Not treated	-	7.5 1.1
RENLRIALRY	**B2702**	**11.4 2.6**
YRLAIRLNER	**-**	**12.1 2.8**
renlrialry	**-**	**11.4 4.1**

yrlairlner	**-**	**13.2 2.7**
RVNLRIALRY	**-**	**11.5 0.5**
YRLAIRLNVR	**-**	**12.5 1.6**
rvnlrialry	**-**	**13.1 3.9**
yrlairlnvr	**-**	**12.2 2.9**
NLRIALRYYW	**-**	**11.6 1.3**
RVNLRTALRY	Kk	8.5 0.7
RVDLRTLLRY	Dk	7.0 0.5
RVDKRTLLGY	Kb	7.8 1.0
RVSLRNLLGY	Db	8.0 0.5
RESLRLLRGY	07	7.5 0.7
REDLRTLLRY	B2705	7.7 1.2
ENLRIALR	-	8.5 0.7
renlpialry	-	9.5 2.4
RVNLRTLRRY	E	8.0 0.5
RMNLQTLRGY	G	7.5 0.7

All of the 19 peptides were used as a learning base to define the rational design strategy.

Although this initial set is very small, it did enable an efficient definition of the constraints, as is shown by the initial distinction between active compounds and inactive compounds.

Initially, 27 independent conformation descriptors were calculated for each peptide (see Table II below).

The physicochemical and topological descriptors were generated by the TSAR software V2.31 (Oxford Molecular Group, Oxford, United Kingdom).

These 27 descriptors contain the dipolar moment, calculated based on a completely developed conformation of each peptide. The statistical analysis showed that 14 descriptors were inter-correlated. Consequently, they were not useful for the definition of constraints that make a

differentiation between active and inactive peptides. The other 13 descriptors, independent of the conformation, were used to construct a static filter for screening a virtual combinatorial library as described later.

TABLE II

The 27 topological and physicochemical descriptors used

Property	Nature	Included/excluded
Molar mass	*Physical*	*Excluded*
Ellipsoidal volume	Shape	Included
Molecular volume	Shape	Included
Molar refractivity	Topological	Included
Lipophilia (LogP)	Topological	Included
Kappa 1	*Topological*	*Excluded*
Kappa 2	*Topological*	*Excluded*
Kappa 3	*Topological*	*Excluded*
Kappa Alpha 1	*Topological*	*Excluded*
Kappa Alpha 2	Topological	Included
Kappa Alpha 3	*Topological*	*Excluded*
Flexibility	Topological	Included
Kier Chi V4	Topological	Included
Randic index	*Topological*	*Excluded*
Balaban index	*Topological*	*Included*
Wiener index	*Topological*	*Excluded*
Sum of condition E	*Physical*	*Excluded*
Dipolar moment	Physical	Included
Number of C atoms	*Chemical*	*Excluded*
Number of O atoms	*Chemical*	*Included*
Number of N atoms	*Chemical*	*Included*
Number of H atoms	*Chemical*	*Excluded*
Total number of atoms	*Chemical*	*Excluded*
Number of methyl groups	*Chemical*	*Excluded*
Number of ethyl groups	*Chemical*	*Included*
Number of amino groups	*Chemical*	*Excluded*

Number of hydroxyl groups	Chemical	Included

The topological descriptors listed above were used to calculate static filters. Thirteen descriptors (shown in upright characters) supplied information about the characteristics of peptides with an immuno-modulating activity and were used to define constraints to screen a virtual combinatorial library. The other fourteen descriptors proved to be inter-correlated and were excluded from the analysis.

The COMBEX (Synt:em, Nîmes, France) program was used to generate a combinatorial explosion based on a consensus sequence RXXXRXXXXY derived from the learning set, after all active and inactive sequences had been aligned. This sequence left seven positions, represented by "X" in which changes could be made in order to create the library.

All molecules were generated using the SMILES convention and were then converted into a 3D structure using the CORINA software (Oxford Molecular Group, Oxford, United Kingdom).

Initially, natural and non-natural amino acids were included in the substitutes data base. All amino acids were described in terms of physicochemical properties (lipophilia, basicity, aromaticity, etc.) and by topological descriptors (Kier analysis, Balaban index, etc.); using the 35 amino acids, this resulted in 357 combinations, namely 64 billion compounds, which is already too many for the capacity of the computer used.

Some additional data such as the distribution of lipophilia were used, in order to reduce this number. In fact, the structure of the "driver" compound (2702.75-84) had two lipophile domains separated by hydrophile residues. This suggested that this distribution was important for the activity. In order to respect this distribution, the Applicant decided to use the V, I, T, W, L, nL (nL=norleucine) amino acids for each of the seven positions. Thus, the number of compounds in the library was modified to 67 combinations, namely 279 936 compounds.

The corresponding set of properties was calculated "on line" for each structure generated by the COMBEX software, in order to screen the virtual combinatorial library. These properties were analyzed using predefined static filters and dynamic filters based on constraints, and only compounds that satisfied all constraints were selected.

Static filters obtained by variation ranges for the selected descriptors were based on all constraints independent of the conformation defined using the learning set of active and inactive compounds. They were designed using the VARIMAP (Synt:em, Nîmes, France) software.

After screening the library of 279 936 compounds with these static filters, 26 peptides were identified that satisfied all constraints. Peptides in this group were studied from the point of view of their conformational spaces (dynamic filter), and five in particular were studied, called RDP1257, RDP1258, RDP1259, RDP1271 and RDP1277.

The flexible nature of the peptides was then analyzed using molecular dynamic (MD) simulations.

The MD simulations of solvated peptides with periodic conditions were then made using the AMBER 4.1 software (Oxford Molecular Group, Oxford, United Kingdom). It used a duration of 1015 ps for each solvated peptide. The dielectric constant was adjusted to be equal to one. The temperature of the system was initially increased gradually from 10 to 300 K over a period of 15 ps. During the simulation, the temperature was kept constant at 300 $\pm$ 10 K, by coupling to an external bath with a relaxation time of 0.1 ps. The selected time "step" was 1 fs. The calculation time was approximately 0.5 h per ps. A cutoff "based on residues" of 10 Å was used for all unrelated interactions. The list of unbonded pairs was updated every 10 fs and the coordinates were collected every ps during the trajectories, which gave a set of 1015 conformations for each trajectory. No constraints were applied to atoms in any of the trajectories, and no "parasite term" was used in the expression for the energy.

Each conformation was represented by a shape descriptor called the 3D autocorrelation vector, hereinafter called "3D-ACV" [Eur. J. Med. Chem. - Chim. Theor., vol. 19,

page 61 (1984)]. A 3D-ACV set was calculated for each conformation series obtained by molecular dynamics (MD) and then processed using statistics with several variables [Trends in QSAR and Molecular Modeling 92 ESCOM Publishers, page 216 (1993)]. This was done in three main steps (i to iii):

(i) For a given conformation of the studied molecule, in this case a peptide, the corresponding 3D-ACV descriptor was calculated as follows. Distances between all pairs of atoms were calculated. The distribution of these distances was a vector in which each bin was the sum of pairs of atoms in a specific inter-atomic separation range, in other words in which two given atoms were separated by a distance between (r-1) and r Å. In this operation, the step was equal to 1 Å. Obviously, the slightest modification to the conformation of the said molecule will change the distribution of inter-atomic distances, and will therefore modify the 3D-ACV. Consequently, this descriptor is one of the most efficient means of representing conformational shapes of molecules.

(ii) An MD trajectory is the set of conformations adopted by a given molecule during the MD simulation. For each conformation, the corresponding 3D-ACV was calculated "in real time" and memorized. This multiple 3D-ACV that is a function of time is a descriptor of the dynamic activity of the said molecule.

(iii) For the comparison of multiple 3D-ACVs representing the trajectories of the analyzed set of molecules, a principal components analysis (PCA) was carried out on each of these multiple 3D-ACVs. This transformation reduced the dimensions of the data set to a smaller number (in this case to a 2D space) and also projected all trajectories of all molecules onto a plane. Each molecule in this reduced space is represented by a set of points, in other words its conformations throughout the duration of the MD simulation, which represents its conformational space. The molecules could then be compared with each other in terms of conformational spaces. Trajectories and conformational spaces were analyzed using the MULTIDYN software (Synt:em, Nîmes, France).

A dynamic filter was used to screen peptides in terms of conformational spaces, and it appeared that four out of these five peptides occupied the same conformational space, or a very similar space, but that one of them, RDP1277, was different from this point of view.

These five peptides were synthesised and tested in biological survival tests following the allo-heart graft.

At a dose of 10 mg/kg/day, all these peptides expect for RDP1277 were active *in vivo*. No significant prolongation to survival following the graft was observed after the therapy with RDP1277 (MST= 9.0 1.4). However the therapy with all the other peptides did produce a significant prolongation to survival after the graft, varying from 11 to 13 days.

Additional tests with peptide RDP1258 at doses of 1 to 10 mg/kg/day revealed an increased efficienty of the therapy at lower doses. At a does of 1 mg/kg/day, 30% of mice that had received a allo-heart graft and were treated by RDP1258 survived for more than 100 days, whereas the reference peptide 2705.75-84 prolonged survival following the allo-heart graft after therapy at a dose of 80 mg/kg/day, whereas no effect was observed at lower doses. Furthermore, no long term survival following the graft was induced by a therapy using this reference peptide.

Therefore, the strategy according to the invention was used to rationally design several bioactive compounds, for which the activity was demonstrated to exceed the activity of active molecules in the learning base.

Further, any reference herein to prior art is not intended to imply that that prior art forms or formed a part of the common general knowledge.

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A computer-assisted process for providing, identifying and describing molecules with a required activity, involving:

1) providing a learning base composed firstly of active molecular structures known to have the required activity and secondly of inactive molecular structures known not to have this activity or only to express this activity very weakly, using appropriate descriptors;

2) generating static filters representative of the intervals of structural, physicochemical and molecular values that must be satisfied by the molecules if they are to be active, starting from the learning base;

3) screening a combinatorial bank of molecules on the basis of said static filters; and

4) submitting the most promising molecules thus selected to dynamic filtering representative of conformational variation constraints which the molecules must observe in order to be active.

2. A process for providing, identifying and describing molecules with a required activity, involving:

1) creating a learning base starting from a family of known similar molecular structures composed firstly of active molecular structures known to have the required activity and secondly of inactive molecular structures known to not have this activity or to not have this activity at all or only very little, using appropriate descriptors;

2) generating a combinatorial explosion of molecules starting from the learning base;

3) screening the molecules thus generated on the

enrichment base, in terms of molecular diversity provided by each molecule regarding the chosen descriptors;

4) submitting the selected molecules to static filters representative of the structural, physicochemical and molecular variation constraints that must be satisfied by the molecules if they are to be active, and possibly to synthesize and test the selected molecules;

5) submitting the most promising molecules thus selected to dynamic filtering representative of confirmation variation constraints which the molecules must observe in order to be active; and

6) synthesising and testing the molecules thus selected.

3. A process as claimed in claim 2, including if the expected result is not obtained, or if it is only partially obtained, repeating steps 3) and 6) modifying the filters.

4. A process as claimed in either claim 2 or 3, wherein the dynamic filtering is performed using software which characterizes the conformational spaces of arbitrary molecules starting from molecular dynamic trajectories.

5. A process as claimed in any one of the preceding claims, wherein said dynamic filtering uses the shape descriptor made up by the 3D autocorrelation vector (3D-ACV).

6. A process as claimed in claim 5 wherein said dynamic filtering consists of calculating a set of 3D-ACVs for each series of conformations obtained by molecular dynamics and then processing by using statistics with several variables, said processing involving:

(i) calculating, for a given conformation of the

studied molecule, the corresponding 3D-ACV descriptor by determining distances between all pairs of atoms and determining the distribution of these distances in the form of a vector for which each bin is the sum of pairs of atoms in a specific inter-atomic separation range;

(ii) calculating the corresponding multiple 3D-ACV in real time and saving in memory said corresponding multiple 3D-ACV for each conformation;

(iii) carrying out a principal components analysis on each of these multiple 3D-ACVs.

7. A process as claimed in claim 4, wherein the static filter uses physicochemical and topological descriptors, at least partly chosen from:
Molar mass; Ellipsoidal volume; Molecular volume; Molar refractivity; Lipophilia (LogP); Kappa 1; Kappa 2; Kappa 3; Kappa Alpha 1; Kappa Alpha 2; Kappa Alpha 3; Flexibility; Kier Chi V4; Randic index; Balaban index; Wiener index; Sum of condition E; Dipolar moment; Number of C atoms; Number of O atoms; Number of N atoms; Number of H atoms; Total number of atoms; Number of methyl groups; Number of ethyl groups; Number of amino groups; Number of hydroxyl groups.

8. A process as claimed in any one of the preceding claims, for use in pharmacology.

9. A molecule with a required activity, designed by the process according to any one of the preceding claims.

10. A computer-assisted process for providing, identifying and describing molecules with a required activity substantially as herein described with reference to the accompanying drawings.

11. A molecule with a required activity substantially as herein described with reference to the accompanying drawings.

Dated this 14th day of February 2003

SYNT:EM (S.A.)

By their Patent Attorneys

GRIFFITH HACK

Fellows Institute of Patent and

Trade Mark Attorneys of Australia