Anticoagulant rodenticide novel candidates predicted by evolutionary docking

short title: anti-rodenticide evolutionary docking

Coll, J*.
Biotechnology Department. Centro Nacional INIA-CSIC. Madrid, Spain.
* Corresponding author
Emails: julio.coll.m@csic.es; juliocollm@gmail.com
Julio Coll, orcid: 0000-0001-8496-3493

Abstract

To generate high-throughoutput numbers of putative anticoagulant rodenticides, a computational evolutionary method has been applied. In particular, the previously described DataWarrior Build-Evolutionary-Library was optimized to randomly generate made-on-demand brodifacoum-derived children libraries exploring new chemotypes for anticoagulant candidates. The generated children were selected by fitting the brodifacoum-docking cavity of rat-VKORC1 (**v**itamin **K** ep**o**xide **r**eductase **c**omplex **1**) alphafold-modeled using the recent crystallographic human-VKORC1 complex. The non-toxic children predicting the highest affinities (lowest docking-scores) for wild type and resistant rat mutants (Spanish and European) and the lowest affinities for human VKORC1, were consensed by AutoDockVina docking to identify with increased accuracy those anticoagulant candidates predicting rat-VKORC1 nanomolar affinities. The resulting layout library of 150 candidates provided with multi-threshold-adjustable filters, constitute an step forward towards *in silico* fine-tuning elimination of any other off-target biological species. Chemical synthesis pathway predictions and experimental validations remain to be done.

Keywords: VKORC1; brodifacoum; rats; evolutionary libraries; docking; rodenticides; human; mutants

Introduction

World-wide rodenticide resistance among rat populations (*Rattus, sp*) have been increasingly detected, most probably due to the intensive use of anticoagulants [1, 2]. For instance, resistance to warfarin (lethal dosage LD_{50} at 58 mg/Kg) or coumatetralyl (16.5 mg/Kg), and both resistance and bait-death association learning when using stronger difenacoum (1.8 mg/Kg), bromadiolone (1.1 mg/Kg), difethialone (0.56 mg/Kg), flocoumafen (0.46 mg/Kg) or brodifacoum (0.26 mg/Kg)[3, 4] anticoagulants, have been reported. Together with the anticoagulant-derived toxic residues found on several off-target biological species [1, 3, 5-7], which may even include humans, genetic resistances constitute one of the main actual concerns to the use of actual rodenticide anticoagulants. Future anticoagulant rodenticides that could maximize on-target rat lethality while minimizing off-target ecological impacts would be convenient [8].

Most rodenticide anticoagulants are made of antagonists inhibiting the **v**itamin **K** ep**o**xide **r**eductase **c**omplex **1** (VKORC1). VKORC1 enzymes are implicated in the carboxylation of proteins (II, VII, IX, X) required for blood coagulation. VKORC1s are found in most biological species as integral membrane enzymes that maintain the balance between reduced-active Vitamin K and its oxidized-inactivate forms generated after carboxylation.

VKORC1 contains a Vitamin K-binding cavity formed by a top and a transmembrane 4-helix bundle structure [9]. By binding to the VKORC1 Vitamin K-binding cavity, rodenticide anticoagulants inhibit Vitamin K binding, and inactivate carboxylation, causing internal bleeding and animal death.

The recent crystal structure of human-VKORC1 complexes with Vitamin K and/or anticoagulants, including brodifacoum[9] (**Figure 1A,B**) suggested that oxidized Vitamin K binds to the top of the VKORC1 cavity causing an open-to-close Cys51 conformational change that triggers oxred reactions[9]. The top cavity contains Cys132, Cys135, and the binding-sites for the heads of Vitamin K and/or anticoagulants (coumarin/indandione). The VKORC1 binding is stabilized by interactions down into the central cavity of the hydrophobic tails of Vitamin K and/or other anticoagulants including hydrogen bonds to residues Asn80 and Tyr139. Binding induces also the displacement of the Cys43-Cys 51 disulphide bond to the top cavity facilitating interchange of reductions of Cys132/Cys135 and oxidized Vitamin K. After reduction, the top cavity opens up facilitating the exclusion of re-activated Vitamin K to continue with other carboxylation cycle. Actual anticoagulant rodenticides target VKORC1 by mimicking Vitamin K binding at similar binding-cavities and conformational changes.

Most rat resistances to anticoagulants have been associated with genetic mutations (single nucleotide polymorphisms, SNPs) at the *vkorc1* gene. Many of the mutations locate at the top of the rat VKORC1 central cavity [5,10,11,12] (**Figure 1A**). Most abundant European rat resistances, corresponded to Y139F, but other mutations are also abundant, such as Y139/C/S L120Q, L128Q and F63C (first letter corresponds to wild-type amino acid residues, number to the position and last letter to the mutant)[5]. In 12 Spanish regions, our group has identified mutations located at S149I/T in urban (brown rat *Rattus norvegicus*) and E155K/Q in rural (black rat *Rattus rattus*) environments. These Spanish mutations map down the VKORC1 cavity nearby the cell membrane [13] (**Figure 1A, orange spheres**). These mutations cause a model displacement of their nearby α-helix by inducing an small 1-2 Å widening of the central cavity (not shown).

To search for new anticoagulant rodenticides a large number of candidates need to be generated and then screened to maximize on-target rat VKORC1 effects (including those of resistant mutants), while minimize unspecific off-target effects, including humans. Because it is not practical to experimentally screen such large numbers of candidates, alternative computational screening of libraries made of similar anticoagulant molecules were proposed before [8, 14]. Combining the recent availability of crystallized human VKORC1-brodifacoum models[9], the accuracy improvements in modeling protein structures by alphafold[15], the identification of brodifacoum-resistance mutants in spanish rats[13], the recent DataWarrior (DW) Build Evolutionary library possibilities and the previous anticoagulant computationally studies [8, 13, 16], large numbers of brodifacoum-derived candidates were generated.

The here named evolutionary docking belongs to self-supervised generation methodologies that bypass the screening of large chemical data banks to detect new protein ligands. In the present particular case, evolutionary docking was used to augment the repertoire of specific ligands by targeting a defined docking-cavity. Specificity was theoretically increased by adding low molecular weight and hydrophobic criteria aiming to reduce off-target effects. The brodifacoum chemotype was selected as the initial parent molecule from which randomly generate multiple children derivatives, because is among the most potent rodenticides due to its highly hydrophobic tail (LogP = 8.0, penetrating the VKORC1 binding central cavity). Additionally, the brodifacoum chemotype shares hydroxycoumarin heads similarly common to warfarine, coumatetralyl, bromadiolone, difenacoum, difethialone, and flocoumafen. Furthermore, a crystallized brodifacoum-binding model has been recently elucidated in human VKORC1 (**Figure 1**). Nevertheless, because its high hydrophobicity may contribute to unespecificity (by binding to any hydrophobic protein surfaces), and therefore increase off-target effects to any proteins from other biological species, reduction of hydrophobicity was also included among selective criteria during evolution.

To generate libraries-on-demand of brodifacoum-children, the recently released DW Build Evolutionary Library subprogram was chosen, among other possible alternatives being developed using novel transformer technologies[17]. DW randomly generated thousands of brodifacoum children and then selected those best fitting to the brodifacoum binding-cavity. Those children predicting high affinity (low docking-scores) for wild-type and resistant mutant rat-VKORC1 and low affinities for human-VKORC1 were finally displayed on a unique table provided with filter sliders to select the best fitting candidates. AutoDockVina docking affinities were also added to the final table to include consensus affinity values to increase the accuracy of low nanomolar predictions. These strategies and results will facilitate further computational ecological studies to avoid any predictable off-target bioeffects, before computationally predicting chemical synthesis pathway alternatives. The identification of possible chemical synthesis pathways in the future may be facilitated by the advent of modern retrosynthesis predictions[18], including those provided by the IBM RXN server (https://rxn.res.ibm.com/). Experimental validations may follow chemical synthesis.

Computational Methods

VKORC1, and VKORC1-mutant rat models

The amino acid sequences of brown and black rat VKORC1 were downloaded from the nucleic sequences present at the NCBI GenBank (http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucleotide). Since brown (*Rattus norvegicus,* NP_976080) and black rat (*Rattus* rattus, XP_032748408) differed only in VKORC1 amino acid residue 90 (I or L, respectively), with similar hydrophobic side-chains, the brown rat VKORC1 sequence was selected for this study. To be 3D modeled, the spanish rat mutations S149I/T and E155K/Q [13] and the prevalent Y139F resistance mutation, were introduced into the wild type

VKORC1 amino acid sequence. The corresponding protein sequences were submitted to Sokrypton Colab Alphafold. (https://colab.research.google.com /github/sokrypton/ColabFold/ blob/main/AlphaFold2.ipynb)[15]. Similarities in Angstroms Å / number of common α-carbons, were in **R**oot **S**quare **M**ean **D**ifferences (RMSD)(**Figure 1C,D**).

The brodifacoum docking cavity was transferred by alignment of the crystallographic brodifacoum-human VKORC1 model[9] to each of the rat VKORC1 wild-type and mutant models by PyMol. The resulting pdb files contained the VKORC1 ATOMS (TER) line followed by HETATM brodifacoum lines(END). The CONECT lines were removed from the pdb files for the DW subprograms.

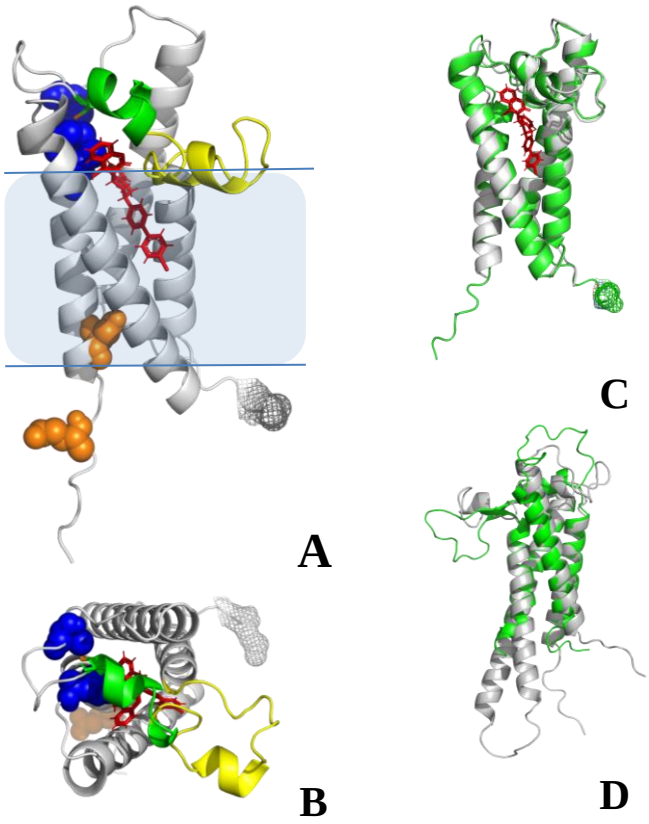

Figure 1
Rat VKORC1-Brodifacoum alphafold complexes and alignments
Rat VKORC1-brodifacoum complex side (A) and top (B) views. Blue horizontal bars and background, cell membrane boundaries. **Gray**, VKORC1 cartoon of the amino acid carbons of the alphafold rat model. **Mesh amino acid**, amino terminal MET1. **Yellow cartoon**, membrane anchor residues 60-80
Green cartoon, cap of the central cavity 52-59 residues. **Blue cartoons**, Cys43-Cys51 and Cys132-Cys135.
Red sticks, brodifacoum
Orange spheres, spanish mutations at residues 149 and 155.
C) Alignment of rat VKORC1 (green) and human VKORC1-brodifacoum[9] (gray) models, RMSD=0.79 Å
D) Alignment of rat VKORC1 (green) and swiss[8] (gray) models, RMSD=1.71 Å

Generation of brodifacoum-children libraries

The "Build Evolutionary Library" subprogram included into DataWarrior generated random brodifacoum-children and selected those fitting the brodifacoum-cavity of the rat alphafold modeled VKORC1. To generate thousands of children, DW randomly adds small molecular modifications to the parent molecule, generating 128 children molecules per generation. The 8-16 best fitting children are then selected for each new generation. Additional fitting criteria can be added such as molecular weight, hydrophobicity (LogP), etc. A number of runs can be repeated starting from the parent molecule which memorize the previously selected children to avoid duplicates. The number of generations stop when reaching a fitness plateau. Memory requirements could be high for a large number of generations/runs, so that monitoring heap memory (Java monitoring console) and RAM memories, were sometimes monitored to reduce program crashes[19, 20].

The raw results were saved as *.dwar files for storage of the complete evolutionary data. The raw children data were then filtered by excluding predicted high and low toxic DW properties (mutagenesis, tumorigenicity, reproductive interference, irritant, and nasty functions) and saved as *.dwar and special *.sdf (vs3) files, maintaining evolutionary information[19, 20] using a home-made DW Macro (**Supplementary information - toxicprediction.dwam**). Toxic-filtered children saved as special *.sdf (vs3) files maintaining the 3D protein cavity docked to children ligand conformers can be opened in PyMol using its split-states command (details at the DataWarrior forum beginning on February 3th, 2023, https://openmolecules.org/forum/index.php?t=msg&th=632&start=0&).

Computational programs employed

The "Build evolutionary library" and "Dock Structures into Docking Cavity" subprograms of DataWarrior (DW) [46] (updated dw550win.zip for Windows) were downloaded (https://openmolecules.org/datawarrior/download.html) following DW guides and our previous work[19, 20]. The AutoDockVina (ADV) program written in Python 3.8 included into PyRx098/PyRx1.0 packages was used to consensus DW docking-scores as described before (https://pyrx.sourceforge.io/). Confirmation of ADV by the recent Vina1.2.3. program was performed with a home-designed colab notebook (**Supplementary Material, Vina123multi.ipynb**). Kcal/mol ADV docking-score outputs were converted to nM Ki by the formula 10^9 *(exp(Kcal/mol/0.592)). MolSoft (ICM Molbrowser vs3.9Win64bit, https://www.molsoft.com/ download.html) was used for manipulating the *.sdf files[19]. Origin (OriginPro 2022, 64 bit, Northampton, MA, USA) (https://www.originlab.com/) was used for calculations and drawings. The predicted structures were visualized in PyRx098/PyRx1.0 (Mayavi), Discover Studio Visualizer v21.1.0.20298 (Dassault Systemes Biovia Corp, 2020, https://discover.3ds.com/discovery-studio-visualizer-download) and PyMOL2.5.3 (https://www.pymol.org/). Hydrophobic and Hydrogen-bonded amino acid interactions in docked complexes were identified by LigPlus2.2.8 (https://www.ebi.ac.uk/thornton-srv/software/LigPlus/download.htm), and visualized in PyMol. A multithreading multi-core i9 (47 CPU) PCSpecialist (AMD Ryzen Threadripper 3960X) provided with 64 Gb of RAM (Corsair Vengeance DDR4 at 3200 MHz, 4 x 16 GB) (https://www.pcspecialist.es/) was used for computations.

Results

The first children libraries obtained from evolutionary-fitting the brodifacoum-rat VKORC1 showed that the brodifacoum-parent predicted more and higher children affinities (lower docking-scores) (**Figure S1, DWR6**) than any of the two other parent fragments derived from brodifacoum containing either its tail or hydroxycoumarin head moieties (**Figure S1, DWR4, DWR5**, respectively). Additional evolutionary-fittings were performed therefore, starting from the whole brodifacoum molecule as the parent to generate brodifacoum-children.

Several brodifacoum-derived libraries were then generated by different criteria. A higher relative weight (x4) was chosen for best fitting to the brodifacoum-cavity to select those with higher affinities. To reduce the numbers of children with undesirable unspecific properties, those with lower hydrophobicity were preferred using logP values <3 but limited by selecting a relative weight of only x1. Preferentially lowering the children hydrophobicity should contribute to reduce hydrophobic-dependent toxicities and unspecific off-target effects, facilitate their handling by increasing aqueous solubilities and most probably, could favor bait taste. To explore and optimize possible affinity effects on the generated children, the "create compound like: approved drugs or natural products" (DW-dependent), the different molecular weight preferences (user-dependent), and the number of runs (user-dependent), were varied.

The number of raw children generated (by introducing random small mutations and atom changes into the brodifacoum parent molecule), varied from 10044 to 26316 children per run (**Table 1**). The percentage of raw children fitting the different criteria mentioned above, varied from 7.1 to 15.9 %, of which 8.8 to 66.6 % were non-toxic. The final numbers of predicted non-toxic children, suggested that large amounts of raw children were required to obtain a few thousands of non-toxic children fitting different molecular weight targets (from 200 to 650 g/mol) (**Table 1**). Most of the generated children reflected the ranges of molecular weights and logP preferences set by the user-defined criteria (not shown). The corresponding ranks profiles were then compared (**Figure 2**).

Table 1
Characteristics of the children generated from brodifacoum with different criteria

Experiment	Chi / Gen	<MW	Runs	Raw Children	Raw Children/Run	Fitting children	%	Non-toxic children	%
B15 +	16	<200	3	33513	11171	5344	**15.9**	3097	**58.0**
DWR10	8	<350	3	37432	12477	2749	**7.3**	1311	**47.7**
DWR11	8	<350	2	26487	13243	1885	**7.1**	904	**48.0**
B2	16	<350	2	44856	22428	6523	**14.5**	1137	**17.4**
B6	16	<400	3	78948	26316	11359	**14.4**	3298	**29.0**
B13 +	16	<400	3	37828	12609	5489	**14.5**	3657	**66.6**
DWR9	8	<500	3	37205	12401	2611	**7.0**	712	**27.3**
B10 +	16	<500	3	38997	12992	5568	**14.3**	3526	**63.3**
B7	16	<522	1	13168	13168	1861	**14.1**	164	**8.8**
DWR8.3 +	8	<650	1	10044	10044	738	**7.3**	391	**53.0**

Initial parent: brodifacoum.sdf. **Target**: binding-cavity of rat VKORC1-brodifacoum.pdb
Common evolutionary criteria and their relative weights(x): fitting to VKORC1-brodifacoum-cavity (x4); < Molecular weight, MW (x2); logP<3 (x1).
Experiment, name of the experiment and color-codes as in **Figure 2.**
Chi / Gen, number of children (Chi) saved between generations of 128 children per generation (Gen)
<MW, children maximal molecular weight criteria during evolution
Runs, number of runs: number of times the evolution restarted from the initial parent to generate further unique children
Raw Children, total number of randomly generated children,
Light blue vertical backgrounds, calculated values
Raw Children / Run, number of raw children generated per run, calculated by: number of raw children / number of runs
Fitting children /%, children fitting the brodifacoum cavity and their % calculated by: 100*fitting children / raw children
Non-toxic children /%, non-toxic children and their % calculated by: 100*non-toxic children / fitting children.
+, DW "create compound like": "approved drugs". **Unlabeled**, DW "create compound like": "natural products".

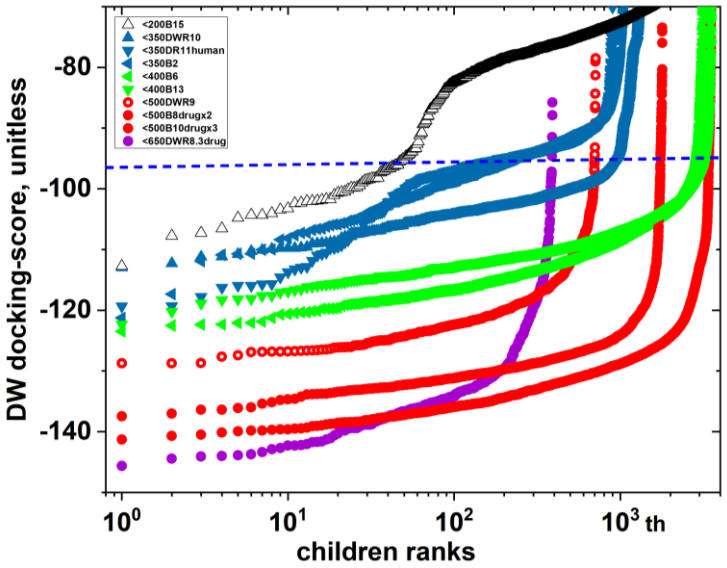

Figure 2
DW docking profiles at different MWs

DW "Build Evolutionary Libraries" were generated selecting different Molecular Weights (MW, g/mol), different number of runs (x) and DW "create compound like": "approved drugs" (drugs) or "natural compounds"(natural). The brodifacoum MW is 522 g/mol.
Blue horizontal dashed line, brodifacoum DW docking-score
Black triangles, MW < 200 g/mol, x3 runs, drugs
Blue triangles, MW < 350 g/mol, x2, x3 runs, natural, rat and human VKORC1
Green triangles, MW < 400 g/mol, x3 runs, drugs and natural
Red spheres, MW < 500 g/mol, x3,x2,x3 runs, natural and drugs
Purple spheres, MW < 650 g/mol, x1 run, drugs

The 128 DW default number of children per generation was used for all the experiments. Increasing from 8 to 16 the number of children passed to the next generation, increased the number of raw children. Numbers >16 were not tested because of the risk of excessively increasing the final numbers of less-fitting children. The number of raw children increased proportionally from 1 to 3 runs per experiment. The number of runs was limited because of increasing their memory resource demands. Memory varied from the most usual heap usage of 5-10 Gb of RAM to the near limit of 50-60 Gb when using molecular weight criteria > 500 g/mol. Since children with > 500 g/mol could also be more difficult-to-be-synthesized, those obtained by higher molecular weights were discarded (i.e., > 650 g/mol). The percentage of fitting children that were non-toxic were between 8.8 to 66.6. Their percentages did not correlated with molecular weights (**Table 1**).

Rank profiles of non-toxic children docking-scores were then comparatively studied for the experiments resumed in **Table 1**. Those children predicting higher affinity ranges (lower docking-scores) than brodifacoum (**Figure 2, blue horizontal dashed line**), were more numerous for molecular weights between 400 to 500 g/mol. The number of non-toxic children reached a few thousands per experiment, 2-3 runs per experiment. There were no high differences in the rank profiles when comparing children obtained by DW "create compound like:" "approved drugs" (**+**) or "natural products" (**Figure 3, green triangles or red spheres**). Only 1.9 or 0.8 % of the non-toxic children structures were repeated between **B6** and **B13** (< 400 g/mol) or among **DWR9, B8** and **B10** (< 500 g/mol) libraries, respectively. Therefore, the non-toxic children predicted using MW criteria < 400 (**B6, B13**) and < 500 g/mol (**DWR9, B8, B10**) were pooled and laydown into an unique table. The removal of the pooled non-toxic children predicting docking-scores > -90 (low affinities), exceptionally low molecular weights < 250 g/mol, and structural duplicates, resulted in 12037 non-toxic children (named library A). Library A contained non-toxic children with DW docking-scores between -90 to -141, molecular weights from 295 to 516 g/mol, and logP from -2.1 to 7.4.

Next, the percentage of children of library A predicting undesirable high affinities to human VKORC1 were investigated. Results showed that only 10.2 % of the 12037 children predicted human-VKORC1 docking-scores equal or lower than those in rats (**Figure S2, blue spheres and blue horizontal continuous line**). These results suggested those children should be excluded for further work because of the risk of being toxic to humans. In this work, we have chosen a rat - human docking-score differential threshold of -25 to be strict in reducing the possibilities of any undesirable interference with human VKORC1 (**Figure S2, blue horizontal dashed line**). Such threshold eliminated 93.7 % of the non-toxic children resulting in library B. Library B contained 747 non-toxic children with maximal affinities to rat VKORC1 (**Figure S2, red points**) and minimal to human VKORC1. Different rat - human thresholds for children elimination could be varied for any future work.

To explore which children of library B retained high-affinity docking to the spanish and European rat resistant mutants, the corresponding mutated amino acids were substituted into the wild-type rat VKORC1 amino acid sequence, alphafold modeled and the docked brodifacoum-cavity transferred from the crystal structure of human VKORC1 complex, as described in methods. DW docking by its "Dock Structures into Docking Cavity" subprogram were then performed by targeting the mutated VKORC1 models with the children of library B. The results were laydown into the same *.dwar file together with those from wild type rat for comparison. Multiple DW filters could be then applied using different thresholds for the rat- and for the human-VKORC1 targets.

To show a prove-of-concept example, library C was generated applying stringent thresholds of > -100 docking-scores to eliminate lower-affinities to rat wild-type and mutants. The opposite filtering direction was applied to discard the children predicting higher affinities corresponding to docking-scores < -100 to human VKORC1 (**Figure 3**). Applying the thresholds mentioned above, the initial library B was downsized from 747 to 150 children (20 % of library B). Any other threshold values may be applied, for instance if adding any new computational scores (i.e., VKORC1 affinities from off-target species) or any experimental *in vitro* or *in vivo* data (i.e., binding to VKORC1 recombinant proteins or rat mortalities after experimental bait supply).

DW cluster analysis of the children structures of library C identified, at least, 4 main scaffolds which contained representative molecules of 3, 4 or 5 molecular rings (**Figure 4**). A graphic representation of the 150 children docked to rat VKORC1, predicted coverage of the top of the VKORC1 cavity. Most of the 150 children were slightly rotated to the opposite part of the brodifacoum hydroxycoumarin head and included similar but smaller hydrophobic tail cavities than brodifacoum (**Figure 5**). The amino acid interactions of the corresponding 4 representative molecular scaffolds, showed maximal numbers of Hydrogen and hydrophobic bonds to amino acids of the rat VKORC1 cavity and numerous newly predicted amino acid contacts (**Table S2, green rectangles**).

AutoDockVina (ADV) docking screening of wild-type rat-VKORC1 added consensus accuracy to the library C predictions. As one example of possible correlations, the 9 ADV best leads to rat VKORC1 were compared to the rat wild-type, mutant and human DW docking-scores. Differences of DW docking-scores were easily seen with human-VKORC1 (**Figure 6, red bars**). Despite the different DW / ADV algorithms correlations between them could be detected (**Figure 6, blue bars**).

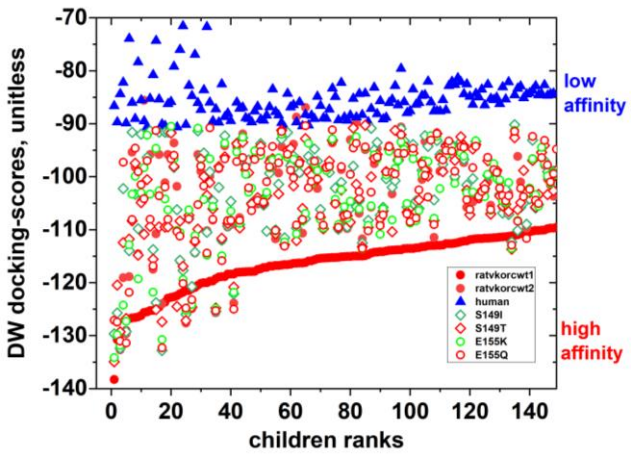

Figure 3
Rank profiles of 150 representatives of the brodifacoum-children library C filtered for high-affinities to rat wild-type VKORC1 and their spanish mutants and low-affinity to human VKORC1
Blue triangles, low affinity children to human VKORC1
Red spheres, high affinity children to wild-type rat VKORC1
Green & red open diamonds, high affinities children to rat VKORC1 S149 resistant mutants
Green & red open spheres, high affinities children to rat VKORC1 E155 resistant mutants

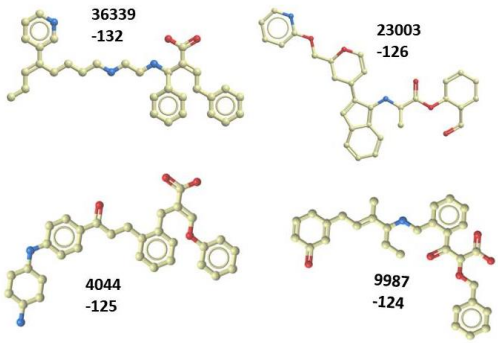

Figure 4
2D structures of representative children library C clusters and their DW docking-scores

https://doi.org/10.26434/chemrxiv-2023-gh4xl-v2 **ORCID:** https://orcid.org/0000-0001-8496-3493

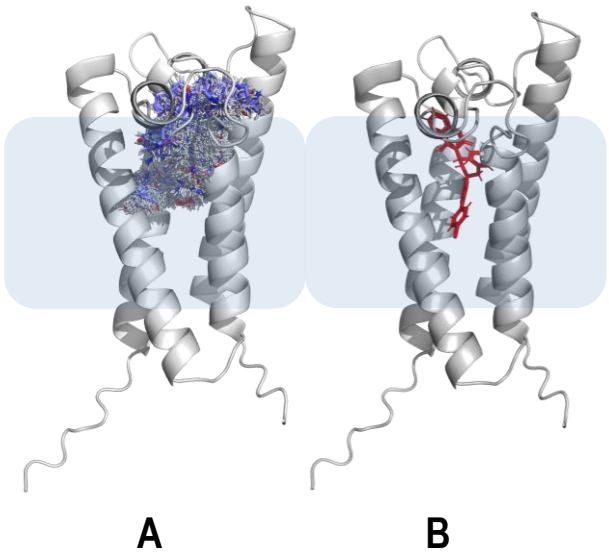

Figure 5
Mapping of 747-children from library B (A) compared to the crystallographic brodifacoum parent (B) DW docked to wild-type rat VKORC1
Blue background, approximated location of cellular membranes
Red stick, crystallographic brodifacoum dock to modeled wild-type rat VKORC1

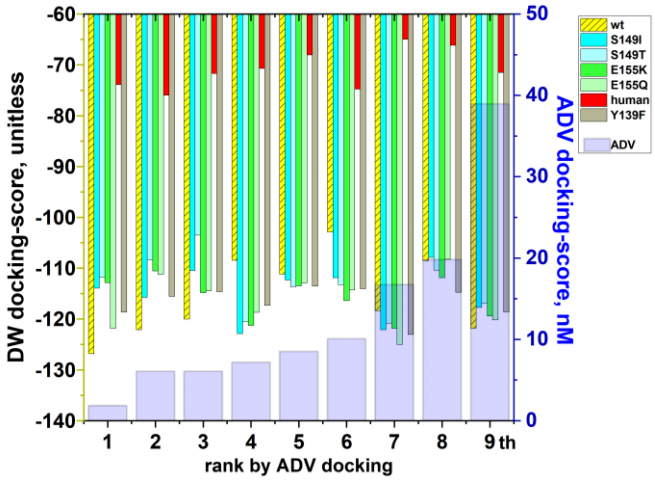

Figure 6
Comparative example of DW and ADV docking results obtained with 9 representative leads
Blue wide bars, mean ADV-docking between PyRx (Vina) and Colab (Vina1.2.3.). The two independent ADV docking-score predictions were very similar.
Red bars, human VKORC1.
Yellow hatched bars, wt rat VKORC1
Blue, light blue, green and light green bars S149I, S149T, E155K, and E155Q mutants, respectively
Gray bars, most abundant European rat VKORC1 mutant Y139F

Discussion

This work was based on the generation of computationally random libraries of thousands of brodifacoum-derivatives (raw children), followed by the selection of those raw children best fitting the brodifacoum-binding cavity of rat VKORC1. The children predictions were more accurate thanks to the recent crystallographic elucidation of brodifacoum and human-VKORC1 complexes.

In addition to fit the rat VKORC1 binding cavity, children molecules were also selected for reduced hydrophobicity (from 8 to 3 logP) and size (from 522 to 400-500 g/mol), to theoretically reduce possible unespecificities. Otherwise, excessive hydrophobicity and/or size could translate into increased off-target biological spreading. Children predicting known toxicities and/or high affinities for human-VKORC1 were also filtered out, serving as examples of how off-target effects could be eliminated from useful predictions. Therefore, the objective of this work was to explore brodifacoum children with increased specificity to reduce off-target effects (in this case only to humans) but conserving high affinities to rat and VKORC1 mutants. In particular, the restricted brodifacoum docking-cavity was employed by DataWarrior (DW) to generate enriched numbers of randomly generated raw children molecules, automatically selecting those with the lowest rat VKORC1 docking-scores (highest affinities), hydrophobicities and molecular weights. Around 12000 unique children molecules could be generated and selected from the vast wide chemical space[21, 22]. Those high numbers of potential anticoagulant rodenticides would had been impossible to reach by any computational screening methods (i.e., AutoDockVina, DEEPScreen, seeSAR, etc), even if applied to search large public chemical libraries containing many millions of compounds (i.e., Mcule, ChemSpace, Zinc, PubChem, Chembl, etc).

For a prove-of concept example showing the flexibility of the approach commented here, the numbers of potentially useful children molecules were downsized to a unique table of 150 candidates, using stringent thresholds maximizing rat wild-type and mutants and minimizing human off-target hits. The method employed here, including the AutoDockVina consensus to confirm low nanomolar affinity predictions, could generated large libraries which could be easily fine tuned to adapt other possible consensus docking data, threshold values and/or starting new evolutionary generations by including any other fitting criteria (i.e, molecular weight, hydrophobicity, number of Nitrogen, Oxygen or rings, etc). Therefore, it should be noted that despite using these powerful approaches, most of the chemical/chemotype space molecular possibilities [21, 22] fitting the rat-VKORC1 cavity still remain to be explored. New algorithms could perhaps be designed to explore further possibilities more exhaustively.

Among the main limitations of the library predictions mentioned above are for example, having explored only one limited VKORC1 docking-cavity, water molecule contributions have not been taken into consideration, and only a limited number of threshold values have been explored. Any exploration of the vast chemical/chemotype space possibilities, including the ones reported here, will always be incomplete. For instance, the explorations did not included any *in vivo* physiological variables (i.e, CYP detoxification). Furthermore, once a few candidates will be more strictly defined (for instance by excluding those off-targeting other biological species in future work), those with easy chemically synthesis will have to be selected for *in vivo* validation. In this aspect and on due course, the identification of possible chemical synthesis pathways might be facilitated by modern retrosynthesis predictions[18], such as those provided by the IBM RXN server (https://rxn.res.ibm.com/). All the above mentioned obstacles still limit the accuracy and possibilities of the predictions made here.

The results described, identified new chemotypes predicting low nanomolar brodifacoum-children with high specificity while conserving its targeting to the main VKORC1-cavity. The results remain to be confirmed after other off-target effects could be minimized by further computational filters and after chemical synthesis [23] would allow experimental validation. Therefore, further work will include selection of those rodenticide candidate molecules with minimal affinity to predator species, chemical synthesis and experimental validation. Perhaps, novel brodifacoum-like second version candidates would be required to be generated to perform those additional selection processes. For instance, possible off-target ecologically important species may include not only well known rat predators but also soil bacteria, fungus and plants, since the VKORC1 enzymes are widely distributed among biological species.

Supporting information

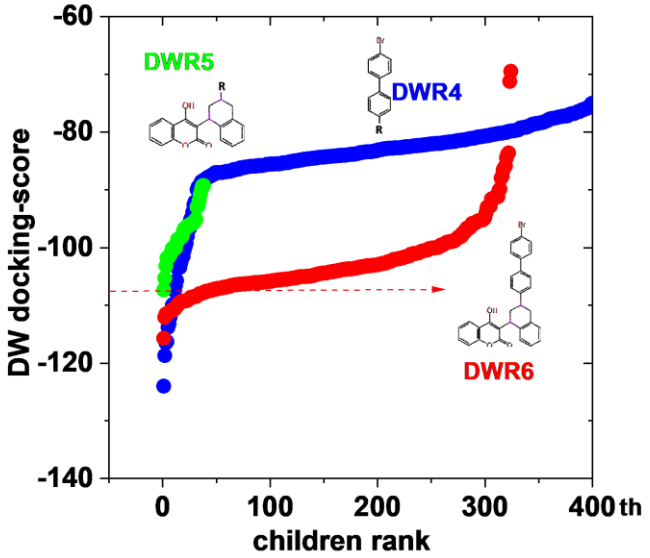

Figure S1
Rank profiles of children from brodifacoum and two brodifacoum derived fragments
The represented children were filtered by known toxicities. The 2D drawings were the structures of the parents used for each of the evolutionary dockings. Criteria used for evolution: 49 generations, 1 run; fitting to the rat VKORC1 brodifacoum-cavity (x4); molecular weight < 650g/mol (x2); logP<3.
Horizontal dashed red line, DW docking-score of brodifacoum.

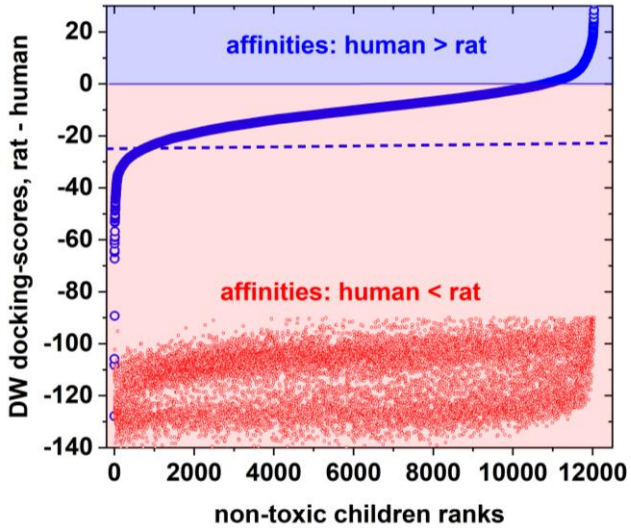

Figure S2
Differential library A of docking-score of non-toxic rank profiles of rat - human VKORC1
Blue open circles, rat-human differential DW docking-scores calculated by the formula, rat docking-score – human docking-score. **Blue horizontal dashed line**, threshold difference of -25 . **Red points**, DW docking-score profile of rat-VKORC1. **Library A**, 12037 non-toxic brodifacoum-derived children

Table S1
Resume of DW docking-scores of representative clustered VKORC1 chemotypes

ID	wt	S149I	S149T	E155K	E155Q	Y139F	human	Vina, nM
36339	-132	-135	-138	-138	-137	-135	-99	90.6
23003	-126	-113	-111	-112	-121	-118	-73	1.8
4044	-125	-109	-115	-115	-115	-116	-99	23.4
9987	-124	-124	-114	-117	-120	-117	-97	90.7

Each of the ID numbers were automatically assigned by the DW generation order during evolution

Table S2
Amino acid residues of rat-VKORC1 predicting contacts with representative non-toxic brodifacoum-derived children leads

position	Aa	Aa	brodifacoum	36339	23003	9987	4044	New sites
19	G	Gly						
22	L	Leu						
23	S	Ser						
26	A	Ala						
30	K	Lys						
54	V	Val						
55	F	Phe					H	
58	R	Arg						
59	W	Trp						
60	G	Gly					H	
61	R	Arg						
62	G	Gly						
63	F	Phe						
77	N	Asn						
78	G	Gly						
79	S	Ser						
80	D	Asp	H					
81	S	Ser			H			
82	I	Ile						
83	F	Phe						
84	G	Gly			H			
85	C	Cys						
87	F	Phe						
88	Y	Tyr						
109	S	Ser						
113	S	Ser						
114	V	Val						
115	A	Ala						
116	G	Gly						
117	S	Ser						
118	L	Leu						
120	L	Leu						
124	L	Leu						
134	V	Val						
135	C	Cys						
138	T	Thr				H		
139	Y	Tyr						
142	N	Asn						

Aa, Amino acid residues of the brodifacoum docking-cavity of wild-type rat VKORC1.
Colored rectangles, amino acid residues predicted by LigPlus
H, Hydrogen bonds predicted by LigPlus between VKORC1 amino acids and non-toxic brodifacoum-derived representative children.
Green, new amino acid contact sites (residues).

Funding: This research was not externally funded .

Competing interests
The author declare no competing interests

Authors' contributions
JC performed and analyzed the dockings, and drafted the manuscript. .

Acknowledgements
Thanks are due to Norwid Behmd of DataWarrior (DW) for valuable contributions to fine tune the toxic filtering macros to apply to the generated children tables from the DW Build Evolutionary Library.

Supplementary Material

-Vina123multi.ipynb. A home-designed Google collaboration notebook (colab web) was employed to confirm PyRx-AutoDockVina docking predictions. The ligands and rat VKORC1 macromolecule pdbqt files and the Vina configuration conf.txt file (including VKORC1 center and docking grid sizes in Angstroms) were prepared for Vina123 docking by the PyRx1.0 (Babel) package, as described before[19,20]. The ligands (*.pdbqt) and rat VKORC1.pdbqt (macromolecule) files were uploaded to a Google drive and docked with the GPU hardware accelerator.

- toxicprediction.dwam. A DW macro dwam file to label and eliminate toxic children molecules generated by the DW Build Evolutionary Library. The macro uses *.sdf or *.dwar files as inputs, user-renamed the input *.dwar file and renamed and saved the corresponding toxic-free *.dwar and toxic-labeled *.sdf files.

-**150B17+adv.dwar**. DW table containing 150 selected putative anticoagulant rodenticide brodifacoum-derived-children. It was provided with threshold slider-filters to their DW docking-scores to VKORC1 (rat, spanish/european mutants, human) and ADV docking-scores to rat VKORC1. The DW table included molecular weights and clogP properties of the putative anticoagulant rodenticides. By up-down moving the slider-filters at the table's right, the best fitting children to particular threshold combinations could be selected. The dwar file can be opened by downloading DataWarrior free access at https://openmolecules.org/datawarrior/download.htm

References

[1]Nakayama, S.M.M., *et al.* A review: poisoning by anticoagulant rodenticides in non-target animals globally.*J Vet Med Sci.* 2019, 81: 298-313 http://dx.doi.org/10.1292/jvms.17-0717.

[2]Nakayama, S.M.M., *et al.* Avian interspecific differences in VKOR activity and inhibition: Insights from amino acid sequence and mRNA expression ratio of VKORC1 and VKORC1L1.*Comp Biochem Physiol C Toxicol Pharmacol.* 2020, 228: 108635 S1532-0456(19)30351-5 [pii], http://dx.doi.org/10.1016/j.cbpc.2019.108635.

[3]Fisher, P., *et al.* Anticoagulant Rodenticides, Islands and Animal Welfare Accountancy.*Animals (Basel).* 2019, 9: ani9110919 [pii], http://dx.doi.org/10.3390/ani9110919.

[4]Tasheva, M. Anticoagulant rodenticides.*WHO IPCS Environmental Health Criteria.* 1995, 175: 124p ISBN9241571756, ISSN 0250-863X.

[5]Buckle, A., *et al.* Resistance testing and the effectiveness of difenacoum against Norway rats (Rattus norvegicus) in a tyrosine139cysteine focus of anticoagulant resistance, Westphalia, Germany.*Pest Manag Sci.* 2013, 69: 233-9 http://dx.doi.org/10.1002/ps.3373.

[6]Chetot, T., *et al.* Vitamin K antagonist rodenticides display different teratogenic activity.*Reprod Toxicol.* 2020, 93: 131-136 S0890-6238(20)30016-2 [pii], http://dx.doi.org/10.1016/j.reprotox.2020.02.003.

[7]Lettoof, D.C., *et al.* Toxic time bombs: Frequent detection of anticoagulant rodenticides in urban reptiles at multiple trophic levels.*Sci Total Environ.* 2020, 724: 138218 S0048-9697(20)31731-9 [pii], http://dx.doi.org/10.1016/j.scitotenv.2020.138218.

[8]Bermejo-Nogales, A., *et al.* Computational ligands to VKORC1s and CYPs. Could they predict new anticoagulant rodenticides?*BioRxiv.* 2021: http://dx.doi.org/10.1101/2021.01.22.426921.

[9]Liu, S., *et al.* Structural basis of antagonizing the vitamin K catalytic cycle for anticoagulation.*Science.* 2021, 371: http://dx.doi.org/10.1126/science.abc5667.

[10]Diaz, J.C., *et al.* Analysis of vkorc1 polymorphisms in Norway rats using the roof rat as outgroup.*BMC Genet.* 2010, 11: 43 1471-2156-11-43 [pii], http://dx.doi.org/10.1186/1471-2156-11-43.

[11]Pelz, H.J., *et al.* The genetic basis of resistance to anticoagulants in rodents.*Genetics.* 2005, 170: 1839-47 genetics.104.040360 [pii], http://dx.doi.org/10.1534/genetics.104.040360.

[12]Czogalla, K.J., *et al.* Structural Modeling Insights into Human VKORC1 Phenotypes.*Nutrients.* 2015, 7: 6837-51 nu7085313 [pii], http://dx.doi.org/10.3390/nu7085313.

[13]Bermejo-Nogales, A., *et al.* VKORC1 single nucleotide polymorphisms in rodents in Spain.*Chemosphere.* 2022, 308: 136021 http://dx.doi.org/10.1016/j.chemosphere.2022.136021.

[14]Coll, J.M. Explorando compuestos rodenticidas con inteligencia artificial.*Jornada TecnicaANECPLA. Avances en elcontrol de roedores y en los estudios de resistencia.* . 2021: On line presentation

[15]Mirdita, M., *et al.* ColabFold: making protein folding accessible to all.*Nat Methods.* 2022, 19: 679-682 http://dx.doi.org/10.1038/s41592-022-01488-1.

[16]Ferencz, L. and Muntean, D.L. Identification of new superwarfarin-type rodenticides by structural similarity. The docking of ligands on the vitamin K epoxide reductase enzyme's active site.*Acta Universitatis Sapientiae Agriculture Environment.* 2015, 7: 108-122 http://dx.doi.org/10.115/ausae-2015-0010.

[17]Grechishnikova, D. Transformer neural network for protein-specific de novo drug generation as a machine translation problem.*Sci Rep.* 2021, 11: 321 http://dx.doi.org/10.1038/s41598-020-79682-4.

[18]Schwaller, P., *et al.* Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy.*Chem Sci.* 2020, 11: 3316-3325 http://dx.doi.org/10.1039/c9sc05704h.

[19]Coll, J.M. Evolutionary-docking targeting bacterial FtsZ.*ChemRxiv.* 2023, https://chemrxiv.org/engage/chemrxiv/article-details/6405c36fcc600523a3bcb679:

[20]Coll, J.M. New star-shaped ligands generated by evolutionary fitting the Omicron spike inner-cavity.*ChemRxiv.* 2023: https://doi.org/10.26434/chemrxiv-2023-v8gqj.

[21]Polishchuk, P.G., *et al.* Estimation of the size of drug-like chemical space based on GDB-17 data.*J Comput Aided Mol Des.* 2013, 27: 675-9 http://dx.doi.org/10.1007/s10822-013-9672-4.

[22]Mroz, A.M., *et al.* Into the Unknown: How Computation Can Help Explore Uncharted Material Space.*J Am Chem Soc.* 2022, 144: 18730-18743 http://dx.doi.org/10.1021/jacs.2c06833.

[23]Azzouzi, M., *et al.* Design, synthesis, and computational studies of novel imidazo[1,2-a]pyrimidine derivatives as potential dual inhibitors of hACE2 and spike protein for blocking SARS-CoV-2 cell entry.*Journal of Molecular Structure.* 2023, 1285: 135525 http://dx.doi.org/10.1016/j.molstruc.2023.135525.