AUTOSTERE : Systematic Search for Scaffold Replacement Opportunities within Structural Databases

Jonathan R. Heal†\*, Joseph M. Sheridan†, Manisha Kulkarni‡, Neill Liptrott§, Deirdre Egan§, Saye Khoo§, Ian Matthews‡, William D.O. Hamilton†

†RxCelerate Ltd, Dorothy Hodgkin Building, Babraham Research Campus, Babraham, Cambridge, CB22 3FH, United Kingdom.

‡ChemOvation Ltd, 49 Arrivato Plaza, Hall Street, Saint Helens. WA10 1GH, United Kingdom

§University of Liverpool, Pharmacology Research Laboratories, Block H, First Floor, 70 Pembroke Place, Liverpool, L69 3GF, United Kingdom

**Abstract:** Medicinal chemists often bias towards working with scaffolds with which previously they have had direct experience and successes. In this way it is often the case that scaffolds which have proven tractable within a research group are ‘re-used’ across multiple and sometimes unrelated drug targets. With this concept in mind we designed a new computer algorithm AUTOSTERE which could systematically assess the opportunities to replace any part of any molecule within an entire database of known ligand structures with a target scaffold and automatically evaluate the potential designs in the context of the original ligand’s protein environment. As such it performs scaffold replacement on an unprecedented scale and suggests new target opportunities for preferred chemistries rather than the conventional reverse situation. The results of this approach for one scaffold, a substituted triazolinone, applied to a set of 10,426 ligand conformations extracted from the PDB are described. This led to the identification of ~600 novel ligands incorporating the triazolinone scaffolds in complex with their predicted drug targets. From these, design examples are provided for HSP-90, cathepsin K and TIE-2 kinase. A further study involved the searching for possible drug targets for unusual pyridopyrimidine cores. This process resulted in the identification of potential novel HIV reverse transcriptase inhibitors which were synthesized and shown to exhibit similar in vitro potencies to marketed compounds. Overall the methodology described provides a powerful new approach to identify new target opportunities for scaffolds of provenance.

**INTRODUCTION**

The term ‘scaffold hopping’ describes process of identifying isofunctional molecular structures with significantly different molecular backbones[[1]](#endnote-1). One way of accomplishing this is to generate an algorithm capable of utilizing the inputs of one or more active ligands to propose new ligand structures related in some way to the original designs but which are predicted to retain or improve activity. Such approaches have been described widely in the literature and include a variety of 2D and 3D methodologies[[2]](#endnote-2). Within the known 3D methods, scaffold replacement approaches generally rely on the user to specify a part of a known molecule which is to be replaced. The algorithms then attempt to connect each of a database of fragments to the exit vectors of the remaining parts of the molecule so as to generate a list of new structures containing alternate connecting fragments for the user specified region of the molecule. The resulting structures can be then minimized to determine the degree of overall shape or feature similarity to the original compound and may also be evaluated for protein target docking. Example programs embodying these concepts include CAVEAT[[3]](#endnote-3), RECORE[[4]](#endnote-4),[[5]](#endnote-5), NEWLEAD[[6]](#endnote-6) and Scaffold Replacement in MOE (CCG Inc). Success with this type of approach have been described in the literature and have been recently reviewed1,[[7]](#endnote-7). Despite successes, the utility of scaffold hopping and in particular scaffold replacement protocols is challenged by a potential lack of synthetic feasibility of the resulting target molecules, which must be assessed individually by a chemist. In addition, medicinal chemists are often searching for solutions to drug design problems by utilizing familiar chemistry and therefore will frequently attempt to design novel structures which incorporate a new chemotype into an existing competitor lead molecule on this basis. Such insights can lead to efficiency in both hit finding and lead optimization phases of the project since experiences of using a scaffold in one project can lead to expeditious use of the same scaffold in the second. This occurs not only because of the practical chemistry knowledge but also because the associated ADMET issues associated with a scaffold have been worked through, albeit in a different structural context. On this basis certain research groups have been successful in applying the principle of scaffold reuse across multiple projects. Such scaffolds are often referred to as ‘privileged’ whose definition in more recent times has been expanded to describe activity at a plethora of target types[[8]](#endnote-8). This observation motivated us to develop an algorithm which could explore the possibilities of scaffold reuse across multiple drug targets. When considering this goal it soon became apparent that whereas in regular scaffold replacement algorithms the user had to select which part of the molecule was the target for replacement, our scheme required that an exhaustive search was undertaken to evaluate replacement possibilities for all parts of the molecule . It also became clear that the application could be aimed at either ligand-only structure collections (such as the Cambridge Structural database, CSD) or more usefully ligand collections in the context of their protein complexed environment (such as the Protein Data Bank, PDB). In the latter case knowledge of protein binding sites could be advantageously utilized to predict which ligands may offer plausible alternates to the source ligand, by means of pose optimization and scoring. Finally, the utility of the output of the algorithm was expected to be highly dependent upon the ability to control the connection rules between the target scaffold and the residual fragments originating in the candidate ligands, particularly because one reason for selecting a given scaffold would be the knowledge of successful chemical reactions involving that scaffold. The ability to ensure realistic chemistry environments of the connections is thus essential to improve tractability of the generated structures.

The result of this exercise was a new computer program AUTOSTERE. The following sections describe the algorithm in more detail and an evaluation of the performance of the approach using (i) in silico analyses of the results for a selected scaffold (triazolinone) applied across the PDB and (ii) the synthesis and in vitro testing of a second scaffold (pyridopyrimidine) predicted and subsequently confirmed to be a novel non-nucleotide reverse transcriptase inhibitors (NNRTI).

**METHODS**

**Computational Studies** All molecular modelling work was conducted in MOE (CCG Inc). AUTOSTERE workflows were written in SVL language within MOE. Ligand minimizations were conducted using the MMFF94x forcefield with the ReactionField solvation model. A Reaction Field solvation model was used because protein complexes retained explicit water molecules. Proteins were prepared as described in the results section. Hydrogens and partial charges were assigned using the Protonate3D method. The dRINGS descriptor was written in SVL to calculate number of ring systems in the molecule from a graph representation.

**Chemistry**

**Preparation of A001.**

Synthesis of (E)-3-(4-amino-3,5-dimethyl-phenyl)acrylonitrile



In a sealable Pyrex tube with stirrer was added tri-o-tolylphosphine (0.122g, 0.4mmol), palladium (II) acetate (0.09g, 0.2mmol), sodium acetate (0.76g, 9.6mmol), tetrabutyl ammonium chloride (2.22g, 8mmol) and dimethylacetamide (16ml). This mixture was heated to 140 °C and acrylonitrile (0.79g, 12mmol) followed by 2,6-dimethyl-4-bromoaniline (1.6g, 8mmol) were then added to the hot mixture and after introduction of nitrogen the reaction was sealed heated and stirred for 48 hours. After cooling the mixture was filtered through Celite and the solid washed with toluene (100ml). The toluene solution was then washed with water (3 x 10ml) and the aqueous phase was re-extracted with toluene (2 x 10ml). The combined toluene extracts were then washed with brine (2 x 20ml) and the organic phase dried over magnesium sulphate. After filtration and evaporation a brown oil of the crude product was left. This oil was further purified by Flash silica (60g) chromatography eluting with hexane:dichloromethane 90:10 to 50:50. Pooled fractions containing product (TLC, dichloromethane:hexane 60:40, Rf 0.122) left, after evaporation, a yellow oil of the title compound (0.76g, 55% yield >95% pure by LCMS mass found 173.4).

Synthesis of 1H Pyrido[3,2-d]pyrimidine-2,4-dione



To 3-aminopyridine-2-carboxylic acid (5g, 36.2 mmol) was added solid urea (9.02g, 150.4 mmol) and the mixture ground to an intimate brown powder. This mixture was heated to 180 °C, a melt occurring at 140 °C. After 1 hour urea was evident on the flask walls and was melted back into the reaction with a heat gun. Over a further hour the reaction was heated to 200 °C and maintained at this temperature for 1 hour, a brown sold became apparent. LCMS showed the reaction to be complete (mass ion 164) and the reaction was left to cool. The solid was then dissolved in 2M sodium hydroxide (400 ml) and then this solution was neutralised (pH~6) with 1M acetic acid (800 ml). The yellow suspension was filtered and washed with water (3 x 50 ml) and diethyl ether (2 x 50 ml). The water solid (5.57g) was then dried in a vacuum oven overnight at 35 °C over calcium chloride to leave a yellow solid of the title compound (3.89g, 66% yield, mass found 164).

Synthesis of 2,4-dichloropyrido[3,2-d]pyrimidine



To pyrido[3,2-d]pyrimidine-2,4-diol (3.89g, 23.9 mmol) was added N,N- diisopropylethylamine (Hunig’s base, 8.3ml, 47.7 mmol). This was heated to 40 °C for 2 hours in a dry nitrogen atmosphere. After this time, phosphoryl oxychloride (40ml) was added (cautiously) over 20 minutes. The reaction was left stirring at 40 °C and deemed complete after 14 hours (small portion reacted with morpholine and analysed by LCMS). The volatiles were removed in vacuo leaving a dark oil. To this oil was added dichloromethane (400ml) and saturated sodium bicarbonate (450ml) which generated a precipitate which was filtered onto Celite and washed with dichloromethane (160ml) and water (160ml) and added to the original mixed organic aqueous phases. The organic phase was separated and washed with brine and dried over magnesium sulphate. Evaporation left a yellow orange oil of 2,4-dichloropyrido[3,2]pyrimidine (4.56g, 95.8% yield, 98% pure by LCMS, mass found 200.7).

Synthesis of (E)-3-[4-(2-chloro-pyrido[3,2-d]pyrimidin-4-ylamino)-3,5-dimethyl-phenyl]-acrylonitrile



To 2,4-dichloro pyrido[3,2-d]pyrimidine (0.84g, 4,2 mmol) was added dimethylacetamide (24ml) followed by (E)-3-(4-amino-3,5-dimethyl-phenyl) acrylonitrile (0.6g, 3.5 mmol), N,N-dimethylaminopyridine (catalytic) and finally N,N-diisopropylethylamine (0.9ml, 5.25 mmol) and the reaction stirred at 80 °C overnight. The reaction mixture was then evaporated and to the residue was added ethyl acetate (150ml) and water (70ml) and extracted. The ethyl acetate layer was then washed with water (50ml) followed by brine (2 x 50ml) and dried over magnesium sulphate. After filtering and evaporation a dark oil was left (1.26g) containing ~30% product. This was purified by flash silica column chromatography (30g silica), eluting with dichloromethane. Thin layer chromatography in dichloromethane:methanol (95:5%) gave the acrylonitrile starting material Rf=0.667 and product Rf=0.533. Product recovered after fraction pooling was 0.108g (9.2% yield, 98% pure by LCMS).

Synthesis of 4-{4-[-((E)-2-cyano-vinyl)-2,6-dimethyl-phenylamino]-pyrido[3,2-d]pyrimidin-2-ylamino}benzonitrile



To a reacti-vial was added (E)-3-[4-(2-Chloro-pyrido[3,2-d]pyrimidin-4-ylamino)-3,5-dimethyl-phenyl]-acrylonitrile (0.108g, 0.32 mmol), 4-aminobenzonitrile (0.076g, 0.64 mmol) and N,N-diisopropylethylamine (0.112ml, 0.64 mmol) and the mixture stirred and heated to 150 °C overnight. Reaction monitoring showed that 8% of starting material was still present so further amounts of 4-aminobenzonitrile (0.0075g, 0.064 mmol) and N,N-diisopropylethylamine (0.011ml, 0.064 mmol) were added and reaction continued for a further 12 hours. Reaction was then found to be complete so ethyl acetate (6ml) and water (6ml) were added to the vial and transferred to a separating funnel. Further amounts of ethyl acetate (30ml) and water (30ml) were added and after separation the aqueous phase was further extracted with ethyl acetate (30ml). The combined organic phases were then washed with brine and dried with magnesium sulphate and evaporated to leave a brown oil of crude product (0.1497g). The crude product was purified by eluting with dichloromethane through a Isolute silica column (30g, silica) to produce product (0.018g, 14% yield, 95% pure by LCMS, mass found 418).

1H NMR:

JEOL, 500MHz

DMSO d6

δ 2.21 (s, 6H, CH3), δ 6.5 (d, 1H, CH=CH-CN, J = 16.65 Hz), δ 7.5 (m, 4H, Ar), δ 7.67 (d, 1H, CH=CH-CN, J = 16.65 Hz), δ 7.77 (dd, 1H, Ar, J = 4.2 Hz), δ 7.90 (d, 2H, Ar), δ 7.94 (dd, 1H Ar, J = 7.94 Hz, J = 8.55 Hz), δ 8.65 ppm ( d, 1H, Ar), δ 9.73 (s, 1H NH) and δ 10 (s, 1H, NH).

E:Z isomer ratio of 89%:11%.

Synthesis of pyrido[2,3-d]pyrimidine-2,4-diol



A finely ground mixture of 2-aminonicotinic acid (1.95g, 14.12mmol) and urea (3.52g, 58.31mmol) were placed in a round bottomed flask and was stirred and heated initially to 180 °C when the reaction became a melt. The solution was then heated to 200 °C for 1 hour then allowed to cool and dissolved in 2M sodium hydroxide (2 x 25ml) followed by the addition of acetic acid (2 x 100ml) which caused a precipitate to form. The solid was filtered and washed with water (3 x 10ml) and ether (2 x 10ml) and then dried overnight in a vacuum oven at 40 °C. This left a yellow powder of the title compound (2.06g, 90% yield, >95% pure by LCMS, mass found 164.7).

Synthesis of 2,4-dichloropyrido[2,3-d]pyrimidine



To a two necked round bottomed flask was added pyrido[2,3-d]pyrimidine-2,4-diol (2g, 12.3mmol) and the flask was cooled to 0 °C in an ice salt bath. To this was then added diisopropylethylamine (4.7ml, 24.5mmol) and the reaction was stirred for 30 minutes. To this cooled reaction was added phosphoryl oxychloride (20ml) and after stirring for a few minutes was heated to 130 °C for 3 hours under nitrogen. The reaction mixture was diluted with dimethylacetamide (20ml) and then evaporated in vacuo. Water (30ml) was then added and the mixture was filtered through Celite and extracted with dichloromethane (2 x 20ml) and the combined organic phase was washed with water (50ml) and then brine (50ml) and dried over magnesium sulphate. After evaporation the crude product was triturated twice with ether to produce a white solid of the title compound (0.77g, 32% yield, 88% pure by LCMS, mass found 200.4) and was used in the next reaction without any further purification.

Synthesis of (E)-3-[4-(2-chloro-pyrido[2,3-d]pyrimidin-4-ylamino)-3,5-dimethyl-phenyl]-acrylonitrile



2, 4-Dichloropyrido[2,3-d]pyrimidine (0.27g, 1.36 mmol) was dissolved in DMA (10ml) in a round bottomed flask to which was added (E)-3-(4-amino-3,5-dimethyl-phenyl) acrylonitrile (0.234g, 1.36 mmol) followed by N,N-diisopropylethylamine (0.355ml, 2.04 mmol) and a catalytic amount of N, N-dimethylaminopyridine. The reaction mixture was heated to 60 °C for 5 days. The reaction mixture was diluted with ethyl acetate (50ml) and this solution was washed with water (3 x 30ml) and brine (2 x 30ml). The organic phase was filtered through cotton wool, dried over magnesium sulphate and after evaporation, produced a yellow powder that was further purified by flash silica chromatography (15g, silica). The product was eluted off the column through a gradient of dichloromethane:ethyl acetate (3:1 to 5:1) to produce product (0.06g, 86% pure by LCMS) which was used without further purification.

Synthesis of 4-{4-[-((E)-2-cyano-vinyl)-2,6-dimethyl-phenylamino]-pyrido[2,3-d]pyrimidin-2-ylamino}benzonitrile



(E)-3-[4-(2-Chloro-pyrido[2,3-d]pyrimidin-4-ylamino)-3,5-dimethyl-phenyl]-acrylonitrile (0.06g, 0.18 mmol) was placed in a round bottomed flask and dissolved in dimethylacetamide (1.5ml) followed by the addition of 4-aminobenzonitrile (0.106g, 0.9 mmol) and the reaction was heated to 120 °C overnight. The reaction mixture was cooled and diluted with ethyl acetate (20ml) and then washed with water (3 x 10ml) and brine (2 x 10ml). The organic layer was dried over magnesium sulphate and evaporated to leave a yellow oily solid which was further purified by flash silica chromatography (5g silica) eluting with hexane:ethylacetate (2:1) producing a yellow powder of the title compound (0.016g, 21% yield, 98% pure by LCMS, mass found 418.2).

1H NMR:

JEOL, 400MHz

DMSO d6

δ 2.23 (s, 6H CH3), δ 6.5 (d, 1H, CH=CH-CN, J = 16.6Hz) δ 7.47-7.55 (m, 4H, Ar), 7.65 (d, 1H, CH=CH-CN, J = 16.6Hz), δ 7.77 (dd, 1H, Ar), δ 7.90 (d, 2H, Ar), δ 8.77 (d, 1H, Ar), δ 8.89 (d, 1H, Ar), δ 10.02 and δ 9.73 (2 x s, 1H, 2 x NH).

E:Z isomer ration is 70%:30%.

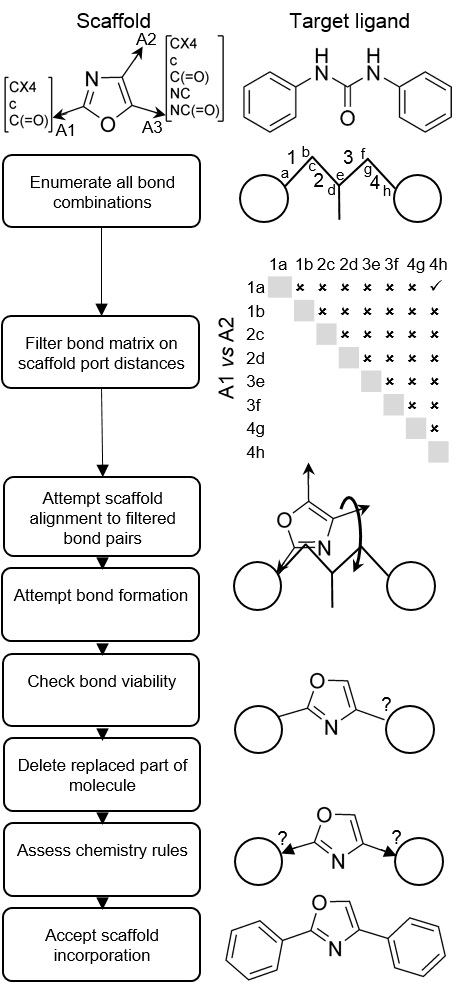
**Biology.** A Roche colourimetric HIV-1 Reverse Transcriptase ELISA was used to assay compound inhibition of HIV-1-RT in vitro. The detection and quantification of synthesized DNA as a parameter for RT activity followed a sandwich ELISA protocol: Biotin-labeled DNA was bound to the surface of microplate (MP) modules that had been precoated with streptavidin. An antibody to digoxigenin, conjugated to peroxidase (anti-DIG-POD), was bound to the digoxigenin-labeled DNA and in the final step, the peroxidase substrate ABTS was added. The peroxidase enzyme catalyzed the cleavage of the substrate, producing a colored reaction product. The absorbance of the samples was be determined using a microplate (ELISA) reader and directly correlates to the level of RT activity in the sample.

A working solution of each compound at 100,000ng/mL using nuclease free water (100μL) was produced. Using a Nunc 96 well plate, a 1:10 serial dilution (100μL), using nuclease free water, of the top concentration vertically down the plate was performed producing an eight point concentration range from 0.01 – 100,000ng/mL. 20μL of each compound was transferred, in triplicate, to a fresh 96 well plate. HIV-1 RT (4ng in 20μL nuclease free water) was added to each well of the 96 well plates containing test compounds but omitted from negative control wells. Nuclease free water (20μL) was substituted for the HIV-1 RT in negative control wells. The Reaction mixture was prepared following the Roche kit instructions and 20μL was then added to each well. Plates were sealed using the plate seals provided and incubated at 37ºC for one hour. Microplate (MP) modules were placed into the MP frame in the correct orientation and following incubation, samples (60μL) were transferred into the wells of the MP modules. Modules were then sealed, covered with foil and incubated at 37ºC for one hour. Following incubation, the solution was completely removed and MP modules rinsed with wash buffer (250μL) five times for 30 seconds ensuring wash buffer was completely removed. Anti-DIG-POD (200μL) was then added to each MP module. Modules were sealed, covered with foil and incubated at 37ºC for one hour. Following incubation, the solution was completely removed and MP modules rinsed with wash buffer (250μL) five times for 30 seconds ensuring wash buffer was completely removed. ABTS substrate solution (200μL) was then added to MP modules and incubated at 15-25ºC until colour development (green) was sufficient for photometric detection (10 minutes). MP modules were shaken (250rpm) then read on a microplate reader measuring the absorbance of the samples at 405nm (reference wavelength 490nm)

IC50 values were calculated using Graphpad Prism software (version 5.0). Absorbance values were normalized to negative control (removal of background absorbance) and entered into software tables against the concentrations used. Data was log transformed and a sigmoidal does-response curve produced. IC50 values for each compound were produced by Graphpad Prism software.

**RESULTS AND DISCUSSION**

The workflow for the generation of new compounds based upon a target ligand and a query scaffold is summarized in Figure 1. The scheme describes the conceptual process by which scaffold replacement is attempted for a candidate ligand. Before the algorithm can run, the user specifies one or more scaffold structures to be used for replacement attempts across the ligand database to be screened. The replacement algorithm process begins with the assignment of possible fragmentation points on the ligand. This is determined by the rotation properties of the bond. All rotatable bonds are considered initially for possible inclusion in a join to the target scaffold. A matrix is then created containing all such pairs of bonds. These bond pairs are subsequently evaluated for distance and angles based on the exit vectors generated for each bond in both directions, since either atom of the rotatable bond is allowed to form a bond with the target scaffold. Based on this calculation, sets of bond pairs are retained which pass a VIABLE\_BOND filter. This filter is passed for a bond pair when the distance between the exit points of the bonds is less than the longest connection point pair on the scaffold plus a padding value, which defaults to 3Å. It also removes any candidate bond pair which is less than the shortest connection point pair on the scaffold. This step removes rotatable bond pairs as connection point candidates because they are either too close, too distant or the angles fall outside the scope of the designated connection points determined by the scaffold. The default value of 3Å to limit processing of too distant rotatable bonds is a rather crude but effective filter in limiting scaffold connection possibilities. It is expected that a smaller threshold of 2.5Å or even 2Å would provide similar results although this has not been rigourously evaluated to date. The remaining bond pair combinations are then subjected to scaffold connection attempts. This process aligns each connection bond of the target scaffold along the exit vector of the first of a pair of rotatable bond candidates or the source ligand using the atom position of the bonded atom to place the heavy atom of the scaffold. The scaffold is then rotated around this new bond at 5º intervals and the distance between the other connection points on the scaffold and other members of the bond pair list are then recorded. If any of the distances are within a bond-like range the scaffold is then bonded to the correct atom in the second bond pair. The parts of the molecule replaced by the new scaffold are then deleted and the new structure is evaluated initially for internal clashes. Any ligands containing heavy atom clashes are removed from the candidate list, whilst ligands containing hydrogen atom only clashes are retained for further optimization. The resultant ligands containing the new scaffolds are subsequently assessed for programmed bond connection rules.



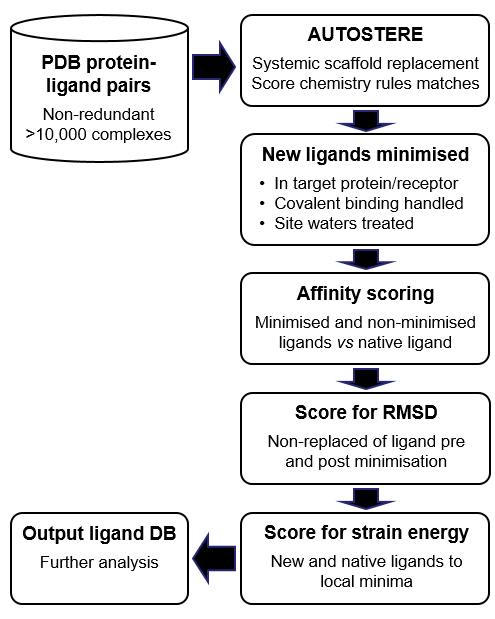
**Figure 1**. Workflow describing scaffold replacement algorithm. All rotatable bonds are considered in the ‘target’ ligand. Bond combinations are filtered further based on calculated distance between the scaffold exit points. Viable bond pairs are then evaluated by rotation of the scaffold while aligned to each of the bond pairs. Once bond viability has been determined the replaced part of the molecule is deleted and chemistry rules are assessed. If the connections meet the chemistry criteria the new scaffold incorporation is accepted.

A target scaffold is defined (Figure 1, scaffold). Each connection point is labeled with a connection type which corresponds to a list of SMARTS environments which this connection type may participate during a connection. When the new ligand is generated using the method described above, the validity of the connection is verified using a SMARTS list for the connection type. Extra rules are encoded to deal with nitrogen forming part of a new bond either as part of the scaffold or as part of the ligand. If a sp2 nitrogen is connected to a sp3 carbon, or if a sp3 nitrogen is connected to sp2 carbon, then the hybridization of the new bonding atoms can be altered to satisfy the environment. A set of SMARTS patterns also determine these rules. SMARTS patterns are defined as a list of text strings and a port number that correspond to the label on the hydrogen atom of the scaffold. Examples of SMARTs lists used for the triazolinone case study is provided in the Supporting Information. This list is read in MOE and scaffold joins which match any of the SMARTS patterns at each port are then recorded in the output database. The ligand is only outputted if all the connections of the scaffold to the parts of the original ligand satisfy specified connection environments. The ligand outputs are stored in a database with flags as to the adherence of the ligand to the chemistry connection environments.

As the AUTOSTERE algorithm is capable of systematic scaffold replacement without the need for the user to specific which parts of the target ligand to replace, it is also interesting to consider that if the method was provided with a database containing a list of new scaffolds to potentially incorporate, an effective single target scaffold hopping tool could be generated. This method would have the advantage over other scaffold hopping approaches of not having to first determine which parts of the ligand to replace, rather, an exhaustive set of new ligands could be generated which would then be triaged by the structural environment of the original ligand, as is provided by the AUTOSTERE workflow methodology.

AUTOSTERE was written in both C++ using modular in house cheminformatic libraries, and the SVL language within the MOE environment[[9]](#endnote-9).

**AUTOSTERE workflows.** The objective of the current research was to systematically apply a scaffold replacement algorithm across a large number of ligands in their bioactive conformations extracted from the PDB such that potentially new target opportunities for that scaffold could be identified. In order to realize this, a workflow was constructed with the purpose of evaluating the viability of the scaffold replaced ligands in the context of the protein structure from which the complexed ligand used as the target was derived. The resulting process for screening a scaffold across the ligands in the PDB and evaluating the resulting new ligand structure against the protein binding sites is shown in Figure 2 below.



**Figure 2**. AUTOSTERE work flow. Each ligand in a database of protein-ligand pairs is the target for scaffold replacement. Successful replacement ligands are individually minimized in the protein structure. Affinity scores are calculated for the original PDB ligand and the new candidate ligands. RMSD values are also generated for the non-replaced parts of the new ligands to determine shift after minimization. Strain energies in final poses are calculated for PDB and new ligands. All of the calculated data is stored in a database along with each ligand.

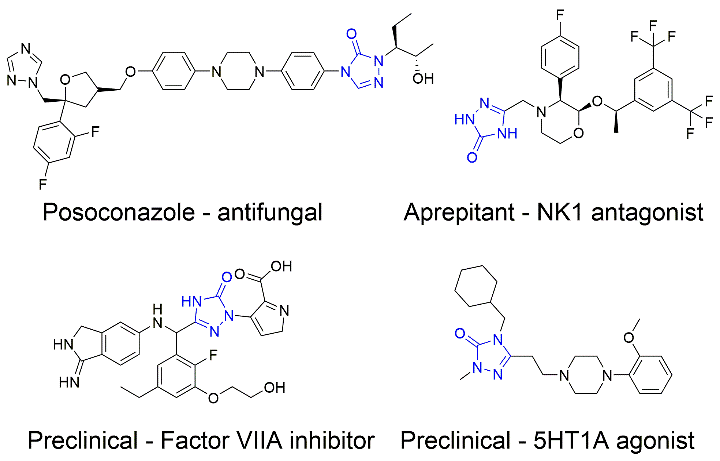
This part of the overall workflow was built within the MOE system (CCG Inc.) using the SVL language. The program initially loads the PDB file from a database containing the target ligand and automatically prepares the binding site for further analysis by identifying the ligand, adding hydrogens, assigning protonation states using the Protonate3Dscript, assigning partial charges (AMBER 99 on the protein, AM1-BCC on the ligand) and determining water molecules in the active site. This final step involves retaining any crystal water molecules within 9Å of any heavy atom of the complexed ligand, having already optimized hydrogen bonding network vis the Protonate3D script. All other water molecules are deleted. Any covalent bonds between the ligand and the protein are recorded for use downstream. The predicted affinity score of the native ligand is then determined using an empirical scoring function PROTOSCORE. This scoring function has been trained on 275 known complexes with Q2 value of 0.82 and has been previously described[[10]](#endnote-10),[[11]](#endnote-11). Each of the scaffold replaced candidate ligands are then loaded in turn into the binding site for further analysis. At this stage site water molecules are assessed for proximity to the new ligand. If the waters are within 1.5Å of the ligand they are removed since the new scaffold may be bulkier than the original ligand captured from the PDB and thus could displace these site water molecules. In addition, if covalent bonds between the native ligand and protein structure exist, these are recreated for the new ligand if those same connecting atoms exist, otherwise the covalent bond is removed. Each candidate ligand is then (i) minimized using MMFF94 forcefield with Born solvation (ii) scored for predicted affinity using PROTOSCORE function and (iii) the RMSD of the parts of the new ligand not containing the scaffold is calculated to the positions of corresponding atoms in the original PDB ligand and (iv) scored for strain energy compared to the lowest energetic minimum. All of this information is stored in a database for further analysis.

**Input Ligand Database Construction** A database of suitable ligands in bioactive conformations extracted from the PDB was required as input to the AUTOSTERE workflow. Initially, a non-redundant set of ligands whose co-ordinates have been extracted from the PDB was downloaded from the Ligand-Expo project[[12]](#endnote-12). The database contained a redundant set of 51,456 non polymer ligands. Redundancy of the ligands structures was determined yielding 6,514 unique ligands in this database. It was important to consider, however, that some of the redundant ligands are captured binding to different protein classes and so the similarity between the proteins was factored into the decision as to which protein-ligand pairs to remove from the initial non redundant set. This consideration led to the use of the SCOP protein classification system to select entries which contained the same ligands but where the protein had a different fold[[13]](#endnote-13). It was expected that the SCOP codes may also prove useful in interpreting the results from scaffold replacement because it was considered likely that a scaffold would prove effective against similar proteins containing similar ligands and the codes would thus provide a way to assess this. The SCOP classification codes were therefore used for clustering entries from the Ligand-Expo database and led to a final set of 10,426 non–redundant protein-ligand pairs for input into the AUTOSTERE workflow.

**Algorithm Validation.** The goal of the AUTOSTERE project was to perform large scale screening for scaffold replacement opportunities. Initially however, it was important to demonstrate that the system could re-discover existing examples of scaffold replacements within ligands represented in the PDB. To investigate this further, a search of the PDB ligand database was conducted to identify pairs of ligands which differed only by one central (non-terminal) group. This method of this analysis was to fingerprint the extracted 10,426 ligand dataset using MACCS keys in MOE and then generate clusters using 85% overlap and threshold values. Ligands which differed from each other only by a single linking group were expected to fall within the same clusters at these threshold values. In total, 14 sets of related PDB structures were identified where ligands differ only by one linking group, once pendant functional groups were removed from the structures. Three of these sets were then selected to test the AUTOSTERE program for its ability to rediscover alternate linking groups observed in these PDB structures. The three test sets were HCV polymerase ligands (PDB files 3BSC, 3C09 and 2FVC), DPP-IV inhibitors (2QTB and 2FJP), and JNK3 inhibitors (3FI2 and 3FI3). In each experiment one ligand was assigned as the ‘target’; an in house fragmentation method was used to derive scaffolds from the other one (or two) ligands before the AUTOSTERE algorithm was subsequently used to screen the target ligand with these scaffolds for suitable replacement opportunities. In each case the generated ligands were minimized in the original receptor of the target ligand and ligand identities, RMSDs and predicted affinity scores were compared allowing a ranking of the solutions.

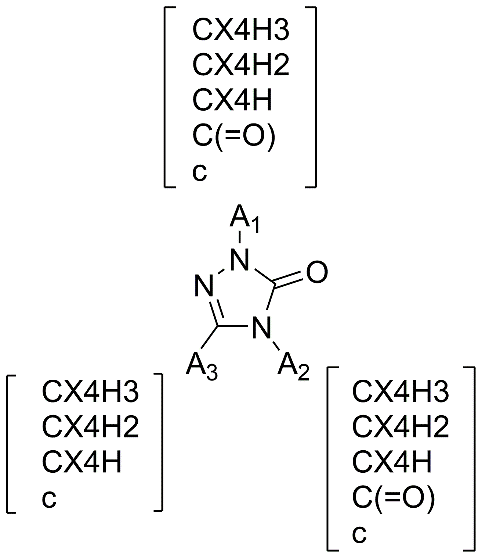
The results showed that in all three cases AUTOSTERE was able to produce a scaffold replacement ligand which was identical to another ligand within the PDB in a very similar conformation to the known bioactive conformation (RMSD range 0.15 - 0.32). The results are shown in more detail in the Supporting Information indicating that the algorithm worked as expected and that subsequent receptor minimization and scoring provides a useful method of determining likely from non-likely ligand replacement suggestions, in which the new ligand conformations are very close (<1Å RMSD for remaining fragments) to those observed in the PDB.

**AUTOSTERE with a triazolinone scaffold**. Having validated that the scaffold replacement procedure operated as expected in test cases, an AUTOSTERE run was conducted across the entire 10,426 non-redundant ligand-protein set using a triazolinone as the screening (replacement) scaffold. This scaffold was selected because it has been observed in marketed compounds[[14]](#endnote-14),[[15]](#endnote-15), such as posoconazole - an antifungal, andother compounds in preclinical or clinical development[[16]](#endnote-16) including selective 5HT1A agonists developed within our group (Figure 3).



**Figure 3**. Triazolinone scaffold (shown in blue) present in marketed (Posoconazole, Aprepitant) and preclinical (Factor VIIA inhibitor,Eisai; 5HT1A agonist, Prosarix) compounds

Not only is the scaffold observed in multiple IND candidates but it has interesting structural properties in its ability to potentially replace a variety of non-cyclic motifs containing H-bond acceptors, such as amides or ureas, potentially increasing rigidity in these and similar polar substructures. It was also considered that this scaffold had a number of connection options thereby increasing the likelihood of identifying matches at different scaffold orientations within the target ligands. The definition of the triazolinone with substitution rules at the 1, 3 and 4 positions is shown in Figure 4. The SMARTS rules associated with this scaffold were that the A1 or A2 hydrogen can be replaced by an aliphatic carbon with either 1, 2 or 3 hydrogens, a carbonyl or aromatic carbon, and the A3 hydrogen can be replaced by an aliphatic carbon with either 1, 2 or 3 hydrogens or aromatic carbon. These substitutions were selected based upon a knowledge of feasible chemistries with triazolinones.

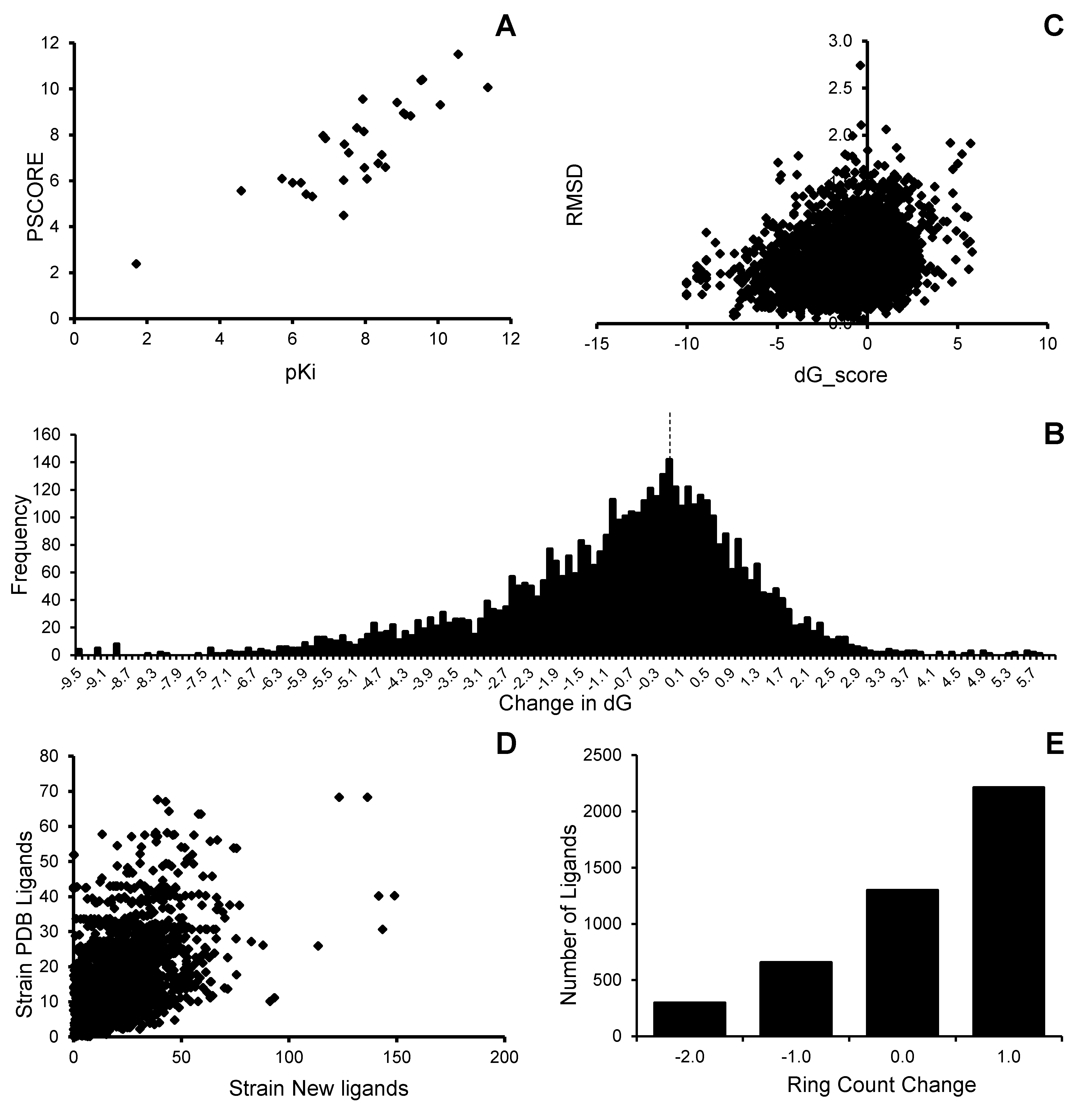


**Figure 4.** Triazolinone scaffold template describing connection chemistry rules. A list of SMARTS patterns define sets for each connection port (replacing a hydrogen) which specifies allowed connections during the scaffold replacement algorithm.

The scaffold was screened for replacement opportunities in the 10,426 ligand set derived from the PDB using the AUTOSTERE program in accordance with the workflow in Figures 1 and 2. The process resulted in the generation of 4,636 ligands containing the new moiety derived from 1,416 PDB entries. This indicates that only 13.6% of the PDB ligands analyzed contained conformations amenable to scaffold replacement with the target group. It is also interesting to note that there is on average 3.3 new ligands per PDB file where scaffold replacement is possible. Closer analysis revealed that the median value is 2 but the range is from 1-19 indicating that some of the more flexible ligands provide high numbers of alternate ligands. This dataset was then evaluated against the protein structures from which the ligands were derived using the second part of the AUTOSTERE workflow. For each ligand the output database contained the predicted affinity score of the native PDB ligand, the predicted affinity score of the new ligand before and after minimization, RMSD of the new ligand compared to the PDB ligand and all site and ligand conformations from the end of the minimization procedures. Based on this data, a number of statistics were derived to summarize different aspects of the results. Since ligands are generated using the bioactive conformation of the PDB ligand but in the absence of the protein, it was expected that the majority of new ligands generated containing the scaffold would either clash with the receptor, leading to lower scoring conformations after minimization, or would introduce strain in the ligand which would lead to predicted affinity scores lower than the PDB complexed ligand. In order to assess the distribution of such scores a dG diff value was determined as dG\_newlig – dG\_PDBlig (both values are predicted pKi values from PROTOSCORE). This metric measures the predicted increase in affinity that results from replacing one part of the PDB ligand with the target scaffold. Therefore positive values indicate a potential increase in affinity from incorporating the scaffold in the design, whereas negative scores indicate a lowering of affinity It should be pointed out however that since the scoring function is only accurate to within ~ +/- 1 log units of affinity, the values have to be considered with this accuracy in mind. Strain energies of the new ligands were also compared to native PDB ligands as a further metric to understand what proportion of generated compounds might be energetically viable.

**Data Analysis.** An initial analysis centered on evaluating the suitability of scoring function for predicting affinity scores of the ligands. As an initial test of the scoring function, the first 30 complexes from the output database were selected and predicted scores were correlated with experimental data. This correlation exhibited an R2 value of 0.64 which was considered acceptable performance of the function based on training set statistics (Figure 5A). None of the 30 complexes were present in the 275 complex training set used to derive the PROTOSCORE scoring function.

Figure 5B shows a histogram of binned dG\_diff values, the predicted increase in affinity of incorporating the scaffold in pKi log units. The data reveals that the average and median values are almost zero, indicating that the most common result is to create a scaffold substitution with little predicted impact on affinity. The data also shows that there are many more examples for where a loss of affinity for scaffold substitution is predicted compared to examples of a predicted improvement of affinity. This is not surprising because the scaffold replacement occurs without knowledge of the protein structure and it would be expected that many of the generated compounds would not optimally interact with the receptor. To analyse the predicted affinity change as a function of RMSD , an RMSD was calculated by comparing positions of common non replaced parts of the PDB ligand and the minimized scaffold-replaced ligand. This data was used to create Figure 5C which shows that there is slight positive correlation (R2 = 0.21) between dG\_diff and RMSD. This demonstrates that favourable scoring scaffold changes tend to alter conformation less relative to lower scoring ligands, although the overall degree of movement is not large since out of 1,626 ligands with a positive dG\_diff value, 1,398 of them have an RMSD < 1 and 502 have an RMSD < 0.5 compared to the native PDB conformation.

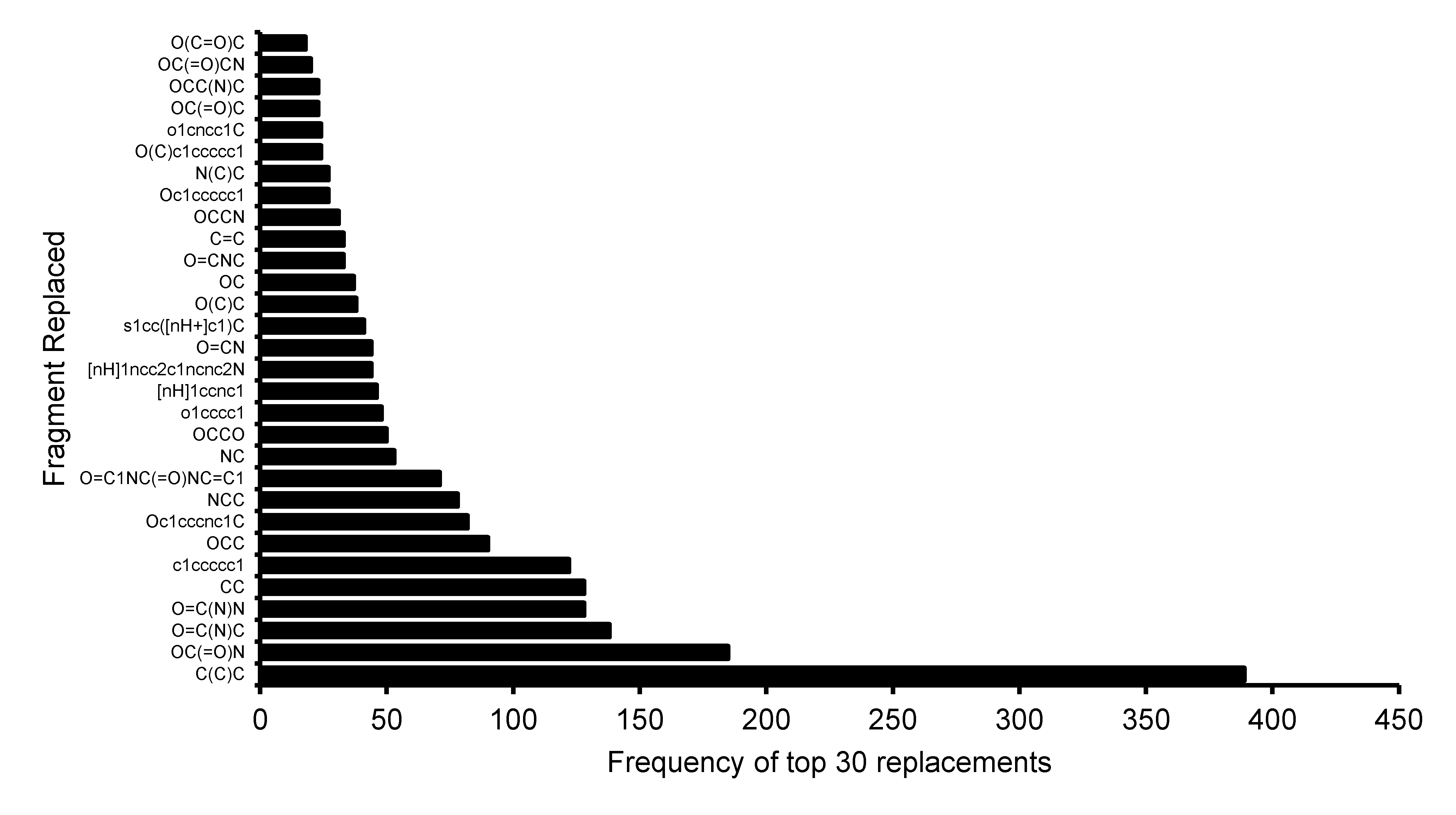


**Figure 5**. (A) Correlation of predicted and experimental pKi values for 30 complexes, (B) Distribution of dG\_Diff values (change in pKi of scaffold ligand – pKi native PDB ligand), (C) Correlation of dG\_diff score with RMSD of non-replaced parts of PDB ligand and scaffold incorporated ligand, (D) correlation of strain energies between the PDB ligands and its scaffold replacement versions, (E) Histogram of the ring count change descriptor (dRINGS).

A further analysis was conducted to evaluate the distribution of strain energies of the generated ligands (Figure 5D). The figure shows the correlation between strain energies (to local minima) between the scaffold incorporated ligands and the corresponding originating PDB ligands. Strain energy for PDB ligands, was obtained using the same minimization protocol as was applied to the generated ligands. Although considerable variation in strain energy has been reported for PDB ligands[[17]](#endnote-17), the data shows that the generated ligands are somewhat more strained than the originating PDB counterparts. This was expected since in the majority of cases the incorporation of the triazolinone scaffold provides a loss of flexibility in the resulting structure, which is likely to induce strain in the context of the protein structure in which it is minimized. However, despite this trend, further analysis reveals that 53% of scaffold incorporated ligands have a strain energy of less than 10kcal/mol. As such it can be concluded that although the current approach is likely to generate more strained ligands than exist in the PDB, AUTOSTERE is able to generate a significant proportion of ligands with low strain energies.

**Analysis of ligand replacements.** It was expected that many of the ligands incorporating a new scaffold would achieve this by replacing a non-ring system. To examine this, a descriptor, dRINGS, was written in SVL to calculate the difference in the number of ring systems in the new ligand compared to the original PDB ligand. The distribution of this descriptor was calculated for the entire data set and is shown in Figure 5E. The data reveals that out of 4,636 successful scaffold replacements, 2,216 involved the replacement of a non-ring fragment of the original molecule. In this respect the algorithm is able to identify many new molecules with less conformational freedom than the PDB ligands. Additionally a further 1,303 generated ligands contain the same number of rings as the source ligands indicating that the triazolinone motif has replaced another ring system.. A smaller pool of 960 ligands contain fewer ring systems after triazolinone incorporation indicating that part of the original molecule has been deleted. This can happen when the part of the ligand being replaced is attached to three groups, two of which meet the chemistry requirements for joining the new scaffold but the third group fails the connection rules and is thus removed from the solution.

The AUTOSTERE program outputs a database of the ligands containing replaced moieties with the target scaffold and also stores the details of the moieties which were replaced from the originating PDB ligand in the output database. This provides the basis for a further analysis into the diversity of molecular groups replaced by the scaffold across a set of results. Such data is expected to be useful in identifying the isosteric diversity of a given scaffold, for example whether the target scaffold almost always replaces the same group. For the current study a histogram was generated for replaced moieties in the originating ligands. From 4636 scaffold replacements across the input database, it was found that 136 different groups were replaced in the input ligands. The histogram showing the most frequent substitutions is shown in Figure 6. The data shows that the most common replacements are of acyclic structures including isopropyl, amides, ureas and carbamates resulting in reduced flexibility of the resultant ligand. The most common ring system to be replaced was phenyl. This methodology therefore is useful in providing an isosteric map for a given functional group, not just from a ligand perspective but also in terms of relevant conformations of such groups for protein binding.



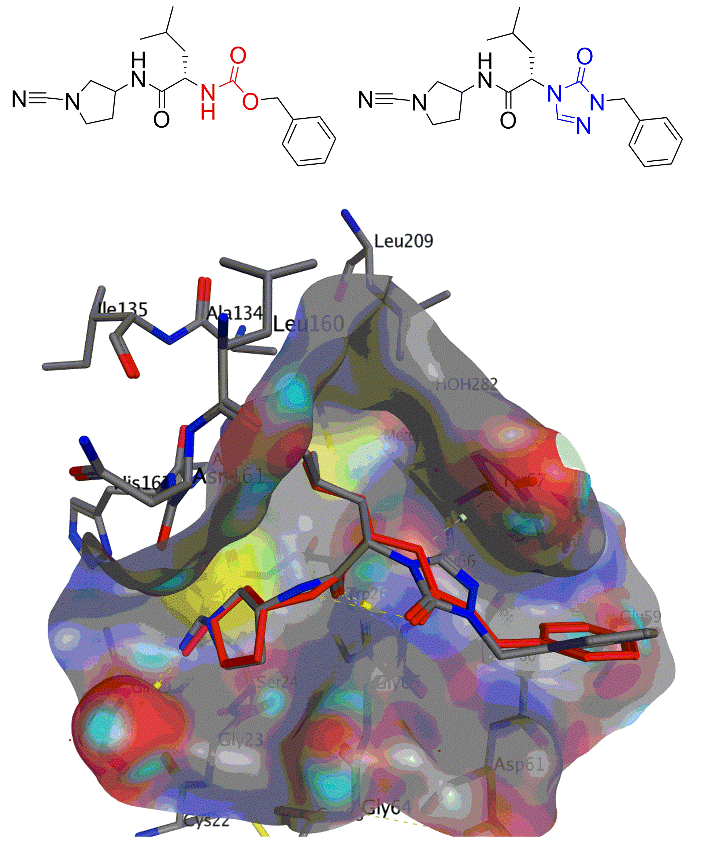
**Figure 6**. Histogram of the top 30 fragments that were replaced (as SMILES) and their frequency of replacement by the triazolinone scaffold.

**AUTOSTERE (triazolinone scaffold) designs.** Having described a number of computational analyses performed across the proteome-wide output, a further set of analyses can be conducted to examine the potential utility of individual results within the dataset. Results were filtered by selecting generated ligands with (i) low strain (<10 kcal/mol) (ii) equivalent or higher predicted affinity score to the native PDB complexed ligand and (iii) at least the same number of ring systems existing in the original PDB ligand. For the triazolinone scaffold query this resulted in the shortlisting of some 600 different targets with novel ligands generated. From within this dataset four examples are provided here in more detail. In the first example an alternate ligand to a known inhibitor of HSP-90 is identified using AUTOSTERE (PDB entry 2BRE). HSP-90 inhibitors exhibit utility as anti-cancer agents[[18]](#endnote-18). Figure 7 shows the chemical structures of the known pyrazole based inhibitor and the AUTOSTERE outputted triazolinone containing structure. The figure details interactions between proposed inhibitor and the protein structure as generated by the program and shows that the computer designed ligand incorporating the triazolinone scaffold generally maintains the binding mode of the original pyrazole containing ligand but provides at least one extra hydrogen bond with a complexed water molecule present in the crystal structure. As such the predicted affinity for the triazolinone is greater than that for the original ligand. The RMSD between PDB ligand and non-replaced parts of the resultant structure is 0.2. This example illustrates that by using the protein structure to score solutions that extra interactions between the protein and solvent may be identified for the scaffold which are absent in the PDB ligand. Having identified that HSP90 was one of the high scoring targets for the triazolinone replacement scheme we subsequently discovered that triazolinones had already been discovered as a potent HSP90 scaffold[[19]](#endnote-19). In particular, a triazolinone based HSP90 inhibitor, similar to the identified design in Figure 8, is currently in phase II. Thus, although the AUTOSTERE design is not novel, it serves as an excellent example of the capability of the algorithm of identifying new opportunities for a selected scaffold.



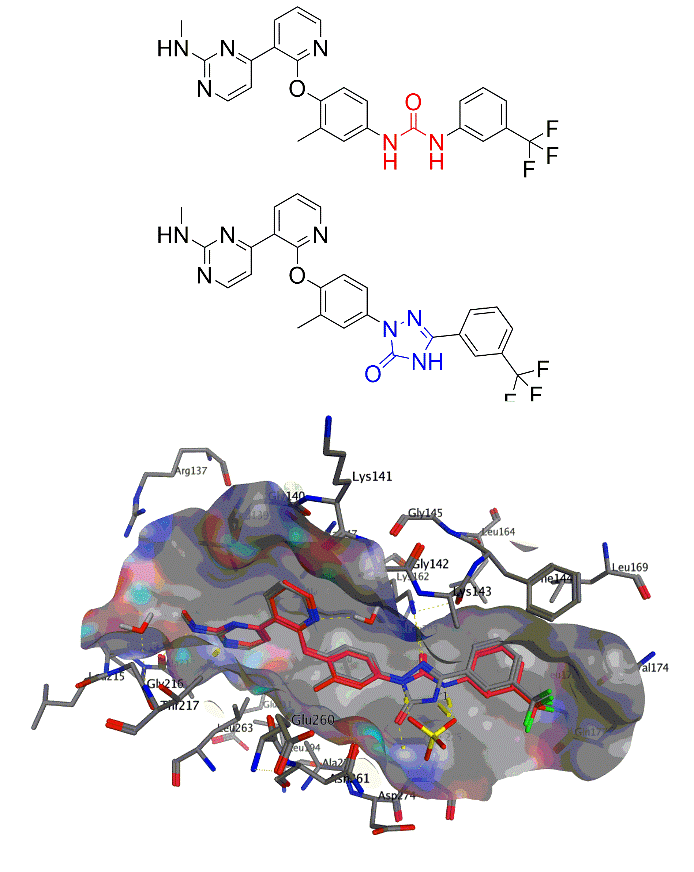
**Figure 7**. HSP-90 example (PDB entry 2BRE). PDB ligand (top-left), AUTOSTERE output ligand (top-right), overlay (bottom) of original compound (red) with AUTOSTERE output ligand (elemental colors).

A second example of AUTOSTERE output is provided by a cathepsin K crystal structure (PDB code 1YK7) in Figure 8. Cathepsin K inhibitors are in the currently in the clinic as potential treatments for osteoporosis[[20]](#endnote-20). In this example the AUTOSTERE outputted ligand contains a triazolinone group replacing a carbamate group of a known inhibitor. The new ligand, despite varying structure at the P3 position is predicted to maintain a similar binding conformation with the receptor to the native ligand (RMSD = 0.18 to common atoms). The outputted ligand retains the same covalent attachment point as in the native PDB ligand. This example demonstrates that the algorithm is capable of successfully handling covalent binding in outputted designs. The resulting ligand appears to be novel as a potential cathepsin K inhibitor.



**Figure 8**. Cathepsin K example. PDB ligand (top-left), AUTOSTERE output ligand (top-right), overlay (bottom) of original compound (red) with the triazolinone replacement compound (elemental).

In the final example a potential novel inhibitor is identified for TIE-2 kinase. Figure 9 shows the structures of the crystalized inhibitor (top left) and the alternative AUTOSTERE outputted compound (top right). The generated ligand contains a triazolinone moiety replacing a carbamate motif in the PDB ligand. The figure shows that the incorporation of the ring system allows for a similar conformation of the pendant phenyl groups but also provides enhanced hydrogen bonding with a complexed phosphate group and also amino acids LYS162 and the backbone amide of PHE275. The generated ligand is thus predicted to represent an alternate scaffold to biaryl ureas to interact with this part of the kinase binding domain.



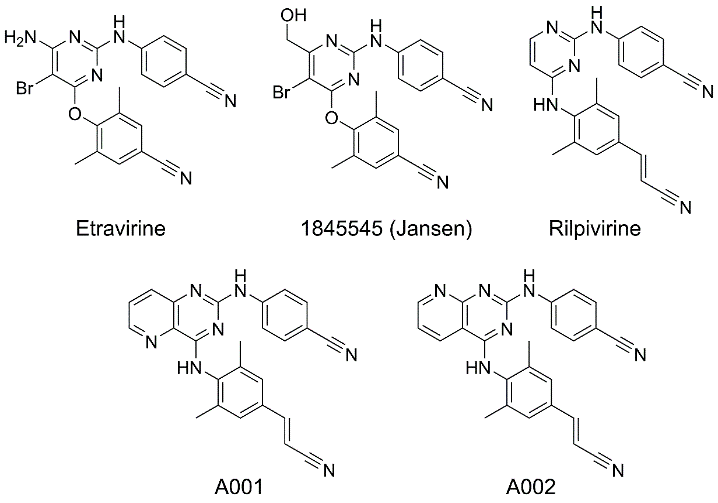
**Figure 9**. Tie-2 kinase example. PDB ligand (top), AUTOSTERE output (middle), overlay (bottom) of original compound (red) with the triazolinone replacement compound (elemental).

**AUTOSTERE (pyridopyrimidine scaffold) designs, synthesis and in vitro testing.** A further AUTOSTERE analysis of the PDB was conducted with two less common pyridopyrimidine scaffolds which we had previous used in a medicinal chemistry program (Figure 10).

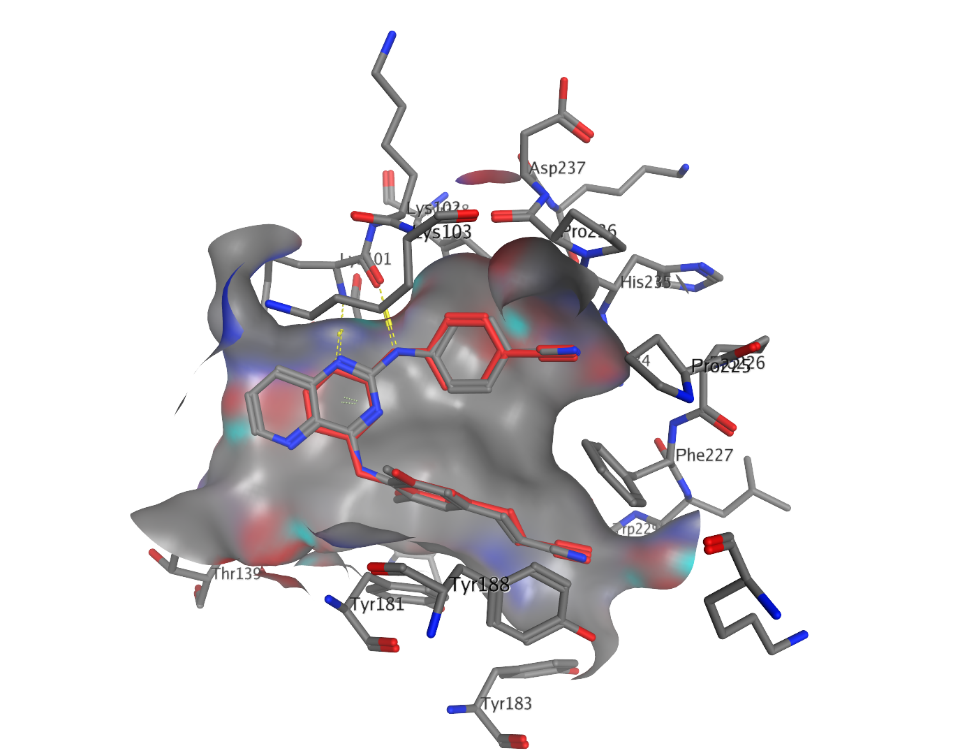


**Figure 10.** Pyridopyrimidine scaffolds. Substitution chemistry was limited to aryl or aliphatic carbon.

The scaffolds have been previously utilized in various settings including inhibitors for mTOR (US20080194546) and PTP-1B (WO2007009911) and we were interested in determining how they may be incorporated as templates in other ligands represented in the PDB. An AUTOSTERE run was conducted in which the allowed chemistry substitutions for the scaffolds were limited to aryl or aliphatic carbon. Out of 10,426 ligand target environments explored, only several hundred solutions were identified in which the replacement of part of the ligand by the pyridopyrimidine scaffold was predicted to have a low strain energy and a similar or increased affinity to the target. It was not possible to conduct an exhaustive assessment of synthetic feasibility of all pyrido pyrimidine compounds generated by the algorithm, but of the ones shortlisted for analysis, at least 50% were considered tractable by a medicinal chemist. The majority of these high ranking solutions targeted kinases, which is somewhat expected given the acceptor/donor features present in the template that can bind to the hinge region of many kinases. However, other highest ranking solutions for both scaffolds were designs predicted to exhibit high affinity for HIV reverse transcriptase targets. For example, PDB files 1SUQ and 2ZD1 both contain NNRTIs complexed with HIV-RT and AUTOSTERE was able to replace the central substituted pyrimidine of the NNRTIs with the pyridopyrimidine scaffolds to predict designs for compounds A001 and A002 shown in Figure 11. Figure 12 shows the predicted binding modes of A001 overlaid with the original PDB ligand rilpivirine in 2ZD1



**Figure 11**. Example NNRTI ligands in the PDB (top), AUTOSTERE outputted pyridopyrimidine NNRTI designs (bottom)



**Figure 12**. HIV-RT in complex with Rilpivirine (PDB 2ZD1, red) overlaid with AUTOSTERE pyridopyrimidine A001 (elemental colours).

It has been described that one of the reasons for selecting a given scaffold for targeted replacement is that it may confer advantageous ADMET properties. In the case of the pyrido pyrimidine scaffold example, that ADMET properties of the predicted and synthesised compounds were calculated to be largely similar to the original compounds present in the PDB. Clearly the potential exists in the future to select scaffolds which have known reduced liability, increased stability and or greater solubility.

In order to test the AUTOSTERE predictions, compounds A001 and A002 were synthesized and tested for their ability to inhibit HIV-RT activity in an *in vitro* assay. For this a Roche colorimetric HIV-1 Reverse Transcriptase ELISA assay was used and the four marketed HIV NNRTI inhibitors were also tested as controls. The two synthesized compounds exhibited potent HIV-RT inhibition, comparable to existing optimized NNRTIs as shown in Table 1.

|  |  |
| --- | --- |
| **Table 1. Compound NNRTI IC50 values** | |
| **Drug/compound** | **IC50 (nM)** |
| Etravirine | 11.81 |
| Nevirapine | 686.82 |
| Efavirenz | <0.01 |
| Rilpivirine | 20.89 |
| A001 | 19.81 |
| A002 | 44.58 |

Overall the data shows that novel NNRTI designs were rapidly identified from a search of the protein-ligand structural space using the AUTOSTERE computer algorithm and an input scaffold for which we were looking to assign new target opportunities. Based on this result and the wider *in silico* results obtained for the triazolinone template search, we expect that this approach has utility in guiding the re-use of scaffolds of provenance across a diversity of drug targets. Clearly the choice of scaffold is an important factor in the likely success of such studies. Small rigid scaffolds with at least one acceptor or donor are likely to be versatile scaffolds due to the variety of protein-ligand interactions they can provide. Additionally points of chirality are likely to increase novelty in the outputted designs. It is also possible that scaffolds extracted from known drugs may bias favourable ADMET properties although this, of course also depends on the final context of the new scaffold.

**CONCLUSIONS**

A new computer program AUTOSTERE was developed which provides automatic scaffold replacement on a database of ligand structures. The procedure benefits from the ability to attempt replacement on all parts of the ligand without user intervention and can automatically minimize, score and thus rank solutions within a protein structure and outputted scores can be compared against the source PDB ligands. The purpose of the program is thus to apply known scaffolds with proven tractability across a large range of targets in order to rank possible new opportunities to best utilize those scaffolds. The algorithm was tested by running a triazolinone motif against a database of 10,426 ligands each in bioactive conformation as recorded in the PDB. Analysis of the results shows that scaffold replacement was achieved in 13% of these ligands and that predicted affinity of the resulting ligands was greater than the original ligands in about third of the solutions. Overall, 600 protein-ligand systems were identified in which the scaffold replaced ligand had low strain energy, low RMSD to original ligand and a predicted affinity at least equal to the original ligand. From these results three examples are provided in more detail for HSP90, cathepsin K and TIE-2 kinase in which the target triazolinone scaffold has replaced part of a co-crystalized ligand. These results exemplify the purpose of the approach that (1) a target scaffold with provenance can be incorporated into many new potentially active leads and (2) that using protein structures to automatically rank the solutions with affinity scoring enables the prioritization of ligands which can including affinity enhancing interactions not present in the original ligand. A further study was conducted for a pyrimidopyridine scaffold which when inputted into AUTOSTERE led to its identification as a potential novel NNRTI scaffold. This was borne out when two synthesized compounds exhibited *in vitro* potencies equivalent to the currently marketed compounds on this target. The ability to identify ligands predicted to interact with a multitude of targets containing a specified scaffold of particular interest to a medicinal chemist provides the platform for chemistry and knowledge re-use across multiple projects. This approach may also be useful for designing focused libraries given a scaffold of interest.

This approach could also be applied to a database of ligand structure conformations in the absence of crystallized proteins (for example the Cambridge Structural Database) providing access to a potentially larger range of targets. In this case field based similarity metrics between the original and scaffold incorporated ligands could be used to rank the ligands. This approach is currently being developed and will be the basis for a further report.

**ASSOCIATED CONTENT**

**Supporting Information Available:**Describes the validation of the algorithm against sets of related ligands from the PDB including tables of RMSD of observed to predicted ligands from known scaffold replacements.

Link: The AUTOSTERE algorithm will be made available on the SVL exchange: <https://svl.chemcomp.com/>

**AUTHOR INFORMATION**

**Corresponding Author**

**Jonathan R. Heal** - RxCelerate Ltd, Dorothy Hodgkin Building, Babraham Research Campus, Babraham, Cambridge, CB22 3FH, United Kingdom.Email: [jon@rxcelerate.com](mailto:jon@rxcelerate.com)

**Authors**

**Jonathan R. Heal** - RxCelerate Ltd, Dorothy Hodgkin Building, Babraham Research Campus, Babraham, Cambridge, CB22 3FH, United Kingdom.

**Joseph M. Sheridan** - RxCelerate Ltd, Dorothy Hodgkin Building, Babraham Research Campus, Babraham, Cambridge, CB22 3FH, United Kingdom.

**Manisha Kulkarni** - ChemOvation Ltd, 49 Arrivato Plaza, Hall Street, Saint Helens. WA10 1GH, United Kingdom

**Neill Liptrott** - University of Liverpool, Pharmacology Research Laboratories, Block H, First Floor, 70 Pembroke Place, Liverpool, L69 3GF, United Kingdom

**Deirdre Egan** - University of Liverpool, Pharmacology Research Laboratories, Block H, First Floor, 70 Pembroke Place, Liverpool, L69 3GF, United Kingdom

**Saye Khoo** - University of Liverpool, Pharmacology Research Laboratories, Block H, First Floor, 70 Pembroke Place, Liverpool, L69 3GF, United Kingdom

**Ian Matthews** - ChemOvation Ltd, 49 Arrivato Plaza, Hall Street, Saint Helens. WA10 1GH, United Kingdom

**William D.O. Hamilton** - RxCelerate Ltd, Dorothy Hodgkin Building, Babraham Research Campus, Babraham, Cambridge, CB22 3FH, United Kingdom.

**Notes**

The authors declare the following competing financial interest(s): J.R.H., J.M.S and W.D.O.H. are employees of RxCelerate Ltd (formerly employees of Prosarix Ltd). M.K. and I.M are employees of ChemOvation Ltd and I.M has a financial interest in ChemOvation Ltd.

**ACKNOWLEDGMENTS**

This work was funded by Prosarix Ltd and ChemOvation Ltd.

**For Table of Contents Use Only**

AUTOSTERE : Systematic Search for Scaffold Replacement Opportunities within Structural Databases

Jonathan R. Heal, Joseph M. Sheridan, Manisha Kulkarni, Neill Liptrott, Deirdre Egan, Saye Khoo, Ian Matthews, William D.O. Hamilton



**REFERENCES**

1. Schneider G.; Neidhart W.; Giller T.; Scmid G. "Scaffold-Hopping" by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angew. Chem. Int. Ed* **1999** *38*, 2894-2896. [↑](#endnote-ref-1)
2. Dobson, C. M. Chemical Space and Biology. *Nature* **2004**, *432*, 824-828. [↑](#endnote-ref-2)
3. Lauri, G.; Bartlett, P. A. CAVEAT: A Pprogram to Facilitate the Design of Organic Molecules. *J Comput Aided Mol Des.* **1994***, 8,* 51-66. [↑](#endnote-ref-3)
4. Maass, P.; Schulz-Gasch, T.; Stahl, M.; Rarey, M. Recore: A Fast and Versatile Method for Scaffold Hopping Based on Small Molecule Crystal Structure Conformations. *J. Chem. Inf. Model.* **2007**, *47*, 390-399. [↑](#endnote-ref-4)
5. VanDrie, J. H. ReCore. *J. Am. Chem. Soc.*, **2009**, *131*, 1617–1617. [↑](#endnote-ref-5)
6. Tschinke, V.; Cohen N.C. The NEWLEAD Program: A New Method for the Design of Candidate Structures Pharmacophoric Hypotheses. *J. Med.Chem*. **1993**, *36*, 3863-3870. [↑](#endnote-ref-6)
7. Sun, H.; Tawa, G.; Wallqvist, A. Classification of Scaffold-hopping Approaches. *Drug Discov Today*. **2012,** *17*, 310-324 [↑](#endnote-ref-7)
8. Welsch, M. E.; Snyder, S. A.; Stockwell, B. R. Privileged Scaffolds for Library Design and Drug Discovery. *Curr Opin Chem Biol.* **2010**, *14*, 347-361. [↑](#endnote-ref-8)
9. *Molecular Operating Environment (MOE)*, 2012.10; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2012. [↑](#endnote-ref-9)
10. Haydon, D.J.; Stokes, N.R.; Ure, R.; Galbraith, G.; Bennett, J.M.; Brown, D.R.; Baker, P.J.; Barynin, VV.; Rice, D.W.; Sedelnikova, S.E.; Heal, J.R.; Sheridan, J.M.; Aiwale, S.T.; Chauhan, P.K.; Srivastava, P.; Taneja, A.; Collins, I.; Errington, J.; Czaplewski, L.G. An Inhibitor of FtsZ with Potent and Selective Anti-staphylococcal Activity. *Science* **2008**, *321*, 1673-1675. [↑](#endnote-ref-10)
11. Bhurruth-Alcor, Y.; Røst, T.; Jorgensen, M.R.; Kontogiorgis, C.; Skorve, J.; Cooper, R.G.; Sheridan, J.M.; Hamilton W.D.O.; Heal, J.R.; Berge, R.K.; Miller, A.D. Synthesis of Novel PPARα/γ Dual Agonists as Potential Drugs for the Treatment of the Metabolic Syndrome and Diabetes Type II Designed Using a New De Novo Design Program PROTOBUILD *Org. Biomol. Chem*. **2011**, *9*, 1169-1188. [↑](#endnote-ref-11)
12. RCSB PDB Protein Data Bank – Ligand Expo Home page. <http://ligand-expo.rcsb.org/> (accessed Nov 12, 2018) [↑](#endnote-ref-12)
13. Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C. SCOP: a Structural Classification of Proteins Database for the Investigation of Sequences and Structures. *J. Mol. Biol.* **1995**, *247*, 536-540. [↑](#endnote-ref-13)
14. Schiller, D. S.; Fung, H. B. Posaconazole: an Extended-spectrum Triazole Antifungal Agent. *Clin Ther*. **2007**, *29*, 1862-1886. [↑](#endnote-ref-14)
15. Hargreaves, R.; Ferreira, J. C. A.; Hughes, D.; Brands, J.; Hale, J.; Mattson, B.; Mills, S. Development of Aprepitant, the First Neurokinin-1 Receptor Antagonist for the Prevention of Chemotherapy-induced Nausea and Vomiting. *Ann N Y Acad Sci*. **2011**, *1222*, 40–48. [↑](#endnote-ref-15)
16. Chang, L. L.; Ashton, W. T.; Flanagan, K. L.; Naylor, E. M.; Chakravarty, P. K.; Patchett, A. A.; Greenlee, W. J.; Bendesky, R. J.; Chen, T.-B.; Faust, K. A.; Kling, P. J.; Schaffer, L. W.; Schorn, T. W.; Zingaro, G. J.; Chang, R. S. L.; Lotti, V. J.; Kivlighn, S.D.; Siegl, P.K.S. Triazolinones as Nonpeptide Angiotensin II Antagonists. 2. Discovery of a Potent and Orally Active Triazolinone Acylsulfonamide. *Bioorg. Med.Chem.Lett*. **1994**, *4*, 115-120. [↑](#endnote-ref-16)
17. Perola E, Charifson PS. Conformational Analysis of Drug-like Molecules Bound to Proteins: an Extensive Study of Ligand Reorganization upon Binding*. J Med Chem*. **2004**, *47*, 2499-2510. [↑](#endnote-ref-17)
18. Kim, Y. S.; Alarcon, S.V.; Lee, S.; Lee, M. J.; Giaccone, G.; Neckers, L.; Trepel, J. B. Update on Hsp90 Inhibitors in Clinical Trial. *Curr Top Med Chem* **2009**, *9*, 1479–1492. [↑](#endnote-ref-18)
19. US20050282119 [↑](#endnote-ref-19)
20. Lewiecki, E. M. Odanacatib, a Cathepsin K Inhibitor for the Treatment of Osteoporosis and other Skeletal Disorders Associated with Excessive Bone Remodeling. *IDrugs*. **2009**, *12*, 799-809. [↑](#endnote-ref-20)