ISSN 0976 – 9080

BIOMIRROR

An Open Access Journal

Drug Design of Novel Molecules Using a Bioisosteric and *De Novo* Techniques - A Comparison

[a] M. Schembri*,
[a] C. Shoemake,
[a] M.A Sant-Fournier

[a] Department of Pharmacy, Faculty of Medicine and Surgery, University of Malta, Msida MSD 2080, MALTA.

ARTICLE INFO

Received 05 September 2014
Revised 07 September 2014
Accepted 09 September 2014
Available Online 14 September 2014

Keywords:
Bioisosteric approach,
De novo design,
Ligand Binding Affinity,
Selective Serotonin Reuptake Inhibitors,
Tricyclic Antidepressants

ABSTRACT

Rational drug design is an area of science that evolves continuously in order to answer contemporary demands for a decrease in novel drug discovery turnover time. Multiple drug design modalities exist which may be exploited in response to the parameters of specific drug design projects. Bioisosteric modification of existing molecules and *de novo* design are two such approaches, both of which were employed in parallel in this study which aimed to compare their scope and efficiency using Tricyclic Antidepressants (TCAs) and Selective Serotonin Reuptake Inhibiting (SSRIs) molecules as case studies. Results indicated that bioisosterically modified structures did not have a higher affinity for their cognate receptor when compared to the template structure while the *de novo* design yielded molecules that were markedly different to the template from a structural perspective, and which also bound to the cognate receptor with an affinity superior to that of the template. This study showed therefore that bioisosteric modification is of utility when minor structural variations are considered sufficiently relative to a template molecule, and could consequently be of utility in the acquisition of new patents, in the reduction of toxicity, or in the attainment of improved biological profiles. It indicated furthermore, the role of the *de novo* approach in the successful exploration of novel pharmacophoric space and in the generation of molecular structures with an affinity significantly greater than that of lead molecules for a target receptor.

Introduction:

Novel drug design is a challenging enterprise that is fraught with numerous pitfalls all of which delay the identification of clinically useful molecular structures[1]. Different rational drug design modalities exist, with their respective advantages and disadvantages.

Corresponding author: M. Schembri*

E-mail address: maria.r.schembri.08@um.edu.mt

Citation: M. Schembri* (Drug Design of Novel Molecules Using a Bioisosteric and *De Novo* Techniques - A Comparison) BIOMIRROR: 59-68 / bm- 0507090514

In the fragment based *de novo* design modality, the judicious planting of high affinity fragments at complementary *loci* within a druggable ligand binding pocket, allows their tethering in the simplest possible way such that the propensity to oral bioavailability is maximised, and the time to reach clinical trial phase is consequently shortened [2]. Bioisosteric replacement also seeks to shorten drug discovery turn over time through the replacement of key molecular moieties with others of near equal molecular shapes and volumes and which also have common electronic distributions and physicochemical properties[3]. Two drug classes were taken as case studies – specifically the Selective Serotonin Reuptake Inhibitors (SSRIs) and the Tricyclic Antidepressants (TCAs). The pharmacophoric space probed was that of a series of bacterial transport proteins, and for each transport protein selected, *de novo*, and bioisosteric approaches were employed in parallel and the novel chemical structures generated through each were compared. The ligand binding pockets of the transport proteins were considered interesting from a drug design perspective owing to their reported roles in the absorption, distribution and elimination of endogenously produced and xenobiotic small molecules [4].

Materials and Methods:

Five Protein Data Bank[5] (PDB) crystallographic depositions were selected for this study. Two of these described bacterial transport proteins bound to commonly used SSRIs, while another three described endogenous transporters (two bacterial and one human) bound to TCA small molecules in common use. Specifically, PDB IDs 3GWV[6] and 3GWU[6] describing the bound co-ordinates of the SSRIs fluoxetine and sertraline respectively, both complexed to leucine transporters (LeuT), PDB IDs 2Q72[7] and 2Q6H[7] describing the tricyclic molecules imipramine and clomipramine respectively, both bound to LeuT, and PDB ID 3APV[8] describing amitriptyline bound to the transport protein human alpha1-acid glycoprotein, were recruited for this study.

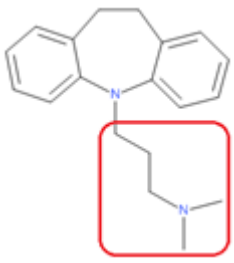
Structure Activity Relationship (SAR) data obtained from the literature identified the moieties critical for binding for each small molecule considered[9,10]. For each crystallographic deposition considered, the respective small molecule ligand was extracted from its cognate ligand binding pocket. Molecular modelling was carried out using SYBYL®-X 1.2[11], SparkV10®[12] was used for bioisosteric molecular generation and *de novo* design was carried out using LigBuilder® v1.2[13]. *In silico* Ligand Binding Affinity (LBA) was calculated using X-SCORE® v1.3[14].

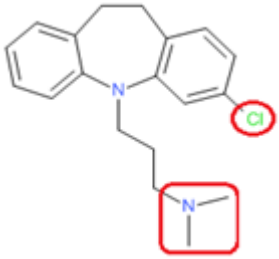
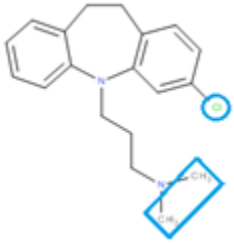
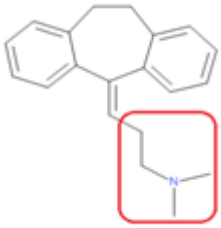
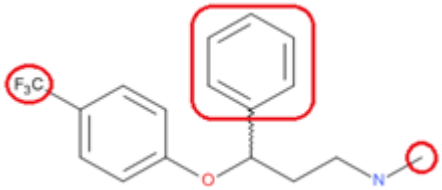
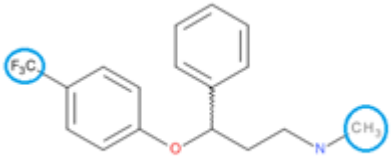
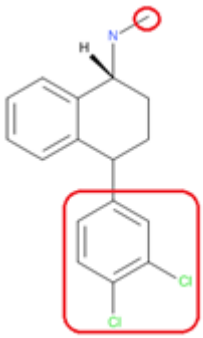
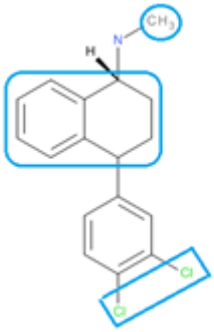
Bioisosteric novel molecular generation involved the identification and modification of *loci* on the template SSRI and TCA molecules other than those constituting the basic pharmacophoric scaffold, and which were considered important for binding. Specifically, molecular modification was user driven in SparkV10®[12] at the *loci* described in Table 1. Molecular modification was carried out on one *locus* at a time. The algorithm embedded in SparkV10®[12] supported the modification of these identified *loci* through the identification of fragments having similar electrostatic and steric properties to the ones singled out for modification from its fragment database. Termination of this process resulted in the generation of novel structures with bioisosterically modified groups at the predesignated molecular *loci*. These novel structures were ranked by the programme according to the Bio-Isostere Factor[15] (BIF) % and Lipinski Rule[16] compliance with the highest BIF%[17] denoting greatest similarity to the parent molecule. The SparkV10®[12] algorithm also predicted which generated bio-isosteric structures would be chemically unstable. These latter were eliminated from the bioisosterically generated pool of molecules. The five highest ranked (according to BIF%[15]) chemically stable structures generated through modification of each *locus* were selected, and modeled in SYBYL®-X 1.2[11] to ensure that these novel structures retained co-ordinates similar to those of the parent ligand. X-SCORE® v1.3[14] was subsequently used in order to measure *in silico* LBA (p*K*d) of the parent small molecule and the five highest ranked bioisosterically modified structures obtained through each successive molecular replacement. The chemical moiety sustaining the highest BIF%[15] for each stepwise bioisosteric substitution was identified and a single molecule bearing all of these substituents was modeled in SYBYL®-X 1.2[11] on the bound co-ordinates of the parent small molecule ligand. This molecule was exported to X-SCORE® v1.3[14] and the *in silico* LBA calculated.

The *de novo* approach was carried out according to a methodology that was essentially opposite to that of the bioisosteric one. Specifically, bioisosteric design involved identifying and modifying those moieties that were considered critical to ligand binding in order to achieve a similar binding modality within a specific Ligand Binding Pocket (LBP). In the *de novo* approach, the moieties previously identified as critical to binding were retained *in situ* as seen in Table 1. These were computationally tethered together according to the genetic algorithm embedded in the *grow* and *link* options of LigBuilder® v1.2[13].

This process involved, for each ligand considered in this study, the creation of a number of *seed* structures (2-3 per small molecule) where the term *seed* implies the collection of fragments critical to ligand binding. These fragments were docked into the LBP at *loci* identical to those described in the crystallographic depositions selected as templates for this study. Seed creation was carried out in SYBYL®-X 1.2[11]. Three dimensional maps, highlighting the polarity of the component *loci* of the LBPs of the receptors considered in this study were generated using the *pocket* algorithm of LigBuilder® v1.2[13].

Table 1: Moieties selected for the TCA and SSRI molecules for bioisosteric replacement (highlighted in red) and the *de novo* approach (highlighted in blue)

Bioisosteric replacement	*de novo* design
Imipramine	
	

Clomipramine	
	
Amitriptyline	
	
Fluoxetine	
	
Sertraline	
	

De novo growth was then sustained, in a user driven manner, within the confines of this delineated space using the *grow* (supports unidirectional molecular growth) and the *link* (allows the conjoining of separate molecular fragments) algorithms of LigBuilder® v1.2[13]. The generated *de novo* structures were then organized into separate families with each family sharing an identical pharmacophoric scaffold, and ranked within each family according to LBA. This procedure was carried out using the *process* algorithm of LigBuilder® v1.2[13], which also gave information, for each novel molecule, regarding its general formula, logP and synthetic feasibility. This information allowed subsequent molecular analysis such that the *de novo* generated structures could be determined to be Lipinski Rule[16] compliant or otherwise.

Results and Discussions:

Imipramine, clomipramine, amitriptyline, fluoxetine and sertraline were considered as template molecules. Each of these was modified at different molecular *loci*. The resultant bioisosterically modified structures (n=20 derived from each parent molecule) were all compliant with Lipinski's Rule of 5[16] and had a LBA (p*K*d) ranging from 6.93 to 5.51.

These results are summarized in Table 2 which shows, for each parent molecule considered, the maximum and minimum LBA (p*K*d) obtained from the respective molecular cohort generated, together with the average recorded value. The LBA (p*K*d) of the parent molecule is included for comparison.

Table 2: Maximum, minimum and average LBA (pKd) obtained for the generated bioisosteres, as calculated in X-SCORE® v1.3

Template molecules		Generated bioisosteres		
Molecule	**LBA (p*k*d)**	**Max LBA (p*k*d)**	**Min LBA (p*k*d)**	**Average LBA (p*k*d)**
Imipramine	6.20	6.53	6.06	6.22
Clomipramine	6.23	6.40	6.07	6.23
Amitriptyline	6.39	6.93	6.29	6.53
Fluoxetine	5.79	6.33	5.51	5.79
Sertraline	6.19	6.34	5.69	6.10

Graph 1 further amplifies this comparative exercise, and shows, of the molecular cohort generated from each parent molecule, the ratio between the bioisosterically generated structures whose LBA (p*K*d) was higher than that of the template resident small molecules and those whose LBA (p*K*d) was lower.

The bioisosterically generated structures resulting from each successive modification exhibiting the highest BIF%[15] scores were merged, when possible, using SYBYL®-X 1.2[11], into a single molecular structure (designated as merged bioisostere in each case).

Graph 1: Graph showing the number of generated bioisosteres that sustained an improvement in LBA (pKd) or otherwise

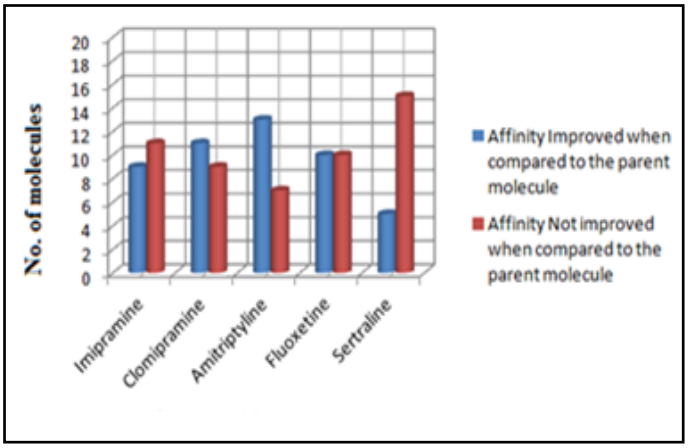
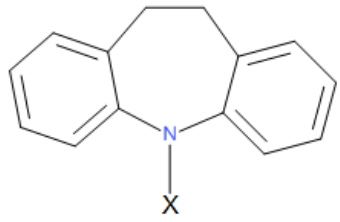
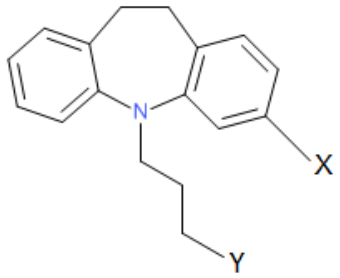
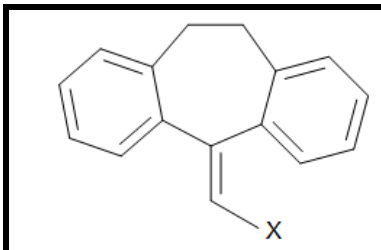
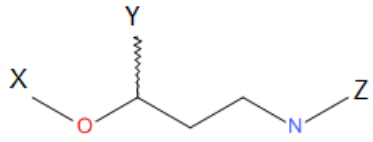
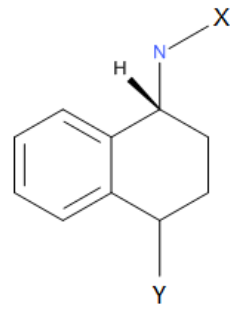


Table 3 summarises the highest BIF%[15] scoring molecules resulting from individual and merged bioisosteric modifications. The *loci* selected for modification are indicated as X, Y and Z. The LBA (p*K*d) score that resulted from each modification is once more included for all the resulting bioisosteres. The LBA (p*K*d) resulting from merging successive modification is highlighted in red.

Table 3: Table showing the LBA (pKd) values of the highest BIF% scoring molecules of individual and merged bioisosteres

		X	Y	Z	BIF %	LBA (p*k*d)
	Imipramine	$-C_3H_6NC_2H_6$	/	/	N/A	6.2
	Bioisostere 1	$-C_6H_{13}$	/	/	92	6.53
	Clomipramine	-Cl	$-NCH_3CH_3$	/	N/A	6.23
	Bioisostere1	-Br	$-NCH_3CH_3$	/	96	6.26
	Bioisostere2	-Cl	$-OCH_3$	/	94	6.14
	Merged bioisostere	-Br	$-OCH_3$	/	N/A	6.17

	Amitriptyline	-$C_2H_4NC_2H_6$	/	/	N/A	6.39
	Bioisostere1	-C_5H_{11}	/	/	98	6.85
	Fluoxetine	-$CF_3C_6H_4$	-C_6H_6	-CH_3	N/A	5.79
	Bioisostere1	-SHC_6H_4	-C_6H_6	-CH_3	94	5.65
	Bioisostere2	-$CF_3C_6H_4$	-C_4H_3S	-CH_3	92	5.54
	Bioisostere3	-$CF_3C_6H_4$	-C_6H_6	-Br	82	5.93
	Merged bioisostere	-SHC_6H_4	-C_4H_3S	-Br	N/A	5.52
	Sertraline	-CH_3	-$C_6H_3Cl_2$	/	N/A	6.19
	Bioisostere1	-Br	-$C_6H_3Cl_2$	/	90	6.34
	Bioisostere2	-CH_3	-C_6H_3ClF	/	93	5.91
	Merged bioisostere	-Br	-C_6H_3ClF	/	N/A	4.88

In the *de novo* design phase of the study, 3 seed structures were generated for imipramine, amitriptyline and sertraline while 2 seed structures were generated for clomipramine and fluoxetine. 600 *in silico* novel structures were generated for imipramine, amitriptyline and sertraline while 400 *in silico* novel structures were identified for clomipramine and fluoxetine. Each molecular cohort generated was segregated by the *process* module of LigBuilder® v1.2[13] into families bearing similar pharmacophoric scaffolds. A total of 45, 32, 52, 42 and 49 families were identified for imipramine, clomipramine, amitriptyline, fluoxetine and sertraline respectively. Within each family, molecules were listed in rank order of decreasing LBA (p*K*d). 10%, 26.5%, 16.2%, 70.5% and 34.5% of the total number of *de novo* generated molecules for imipramine, clomipramine, amitritpyline, fluoxetine and sertraline respectively were identified as being Lipinski Rule[16] compliant.

These results are summarized in Table 4, which shows for each parent molecule considered, the maximum and minimum LBA (p*K*d) obtained from each *de novo* designed molecular cohort generated, together with the average recorded value. The LBA (p*K*d) of the parent molecule is included for comparison.

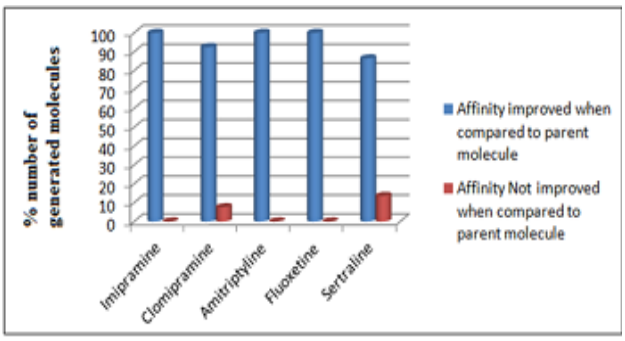
Graph 2 compared, of the molecular cohorts generated from each parent molecule, the ratio between the *in silico de novo* generated structures whose LBA (p*K*d) was higher than that of the template resident small molecule and those whose LBA (p*K*d) was lower.

Table 4: Maximum, minimum and average LBA (pKd) values measured for the generated *de novo* designed structures

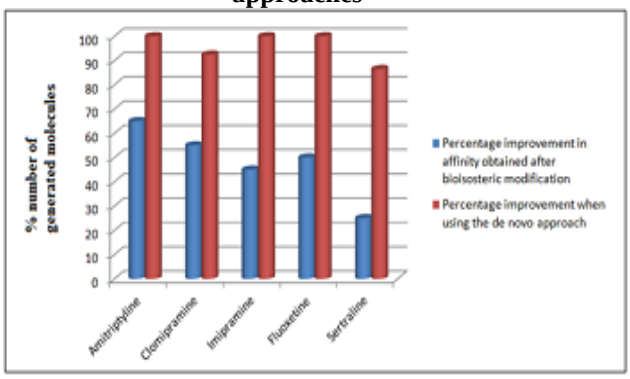
Template molecules		**Generated molecules**		
Molecule	**LBA (p*k*d)**	**Max LBA (p*k*d)**	**Min LBA (p*k*d)**	**Average LBA (p*k*d)**
Imipramine	6.20	9.80	7.08	7.65
Clomipramine	6.23	8.36	6.16	7.06
Amitriptyline	6.39	9.95	7.71	8.58
Fluoxetine	5.79	7.26	5.83	6.22
Sertraline	6.19	8.20	6.02	7.42

Figures 1a-1e are a structural summary of these results. Each parent molecule, imipramine, clomipramine, amitriptyline, fluoxetine and sertraline together with the derived seeds, and the *de novo in silico* designed molecules which complied with Lipinski's rules[16] and exhibited the highest and lowest LBA (p*K*d) respectively are shown.

Graph 2: Graph showing the percentage of molecules that showed an improvement in the LBA (pKd) or otherwise



Graph 3: Graph showing the percentage of molecules that showed an improvement in the LBA (pKd) when using the bioisosteric and *de novo* approaches



Graph 3 compares the percentage improvement relative to each template small molecule with respect to LBA (p*K*d) when bioisosteric modifications and when *de novo* approach were adopted.

Figure 1A: Imipramine template molecule, with its derived seeds and the top 3 *de novo* designed structures having the highest LBA (pKd) for each respective seed

Imipramine	Seed 1	Novel *de novo* designed structures with highest LBA (p*K*d)		
		Family 1	Family 2	Family 3
p*K*d = 6.20	Seed 1	p*K*d = 9.58	p*K*d = 9.35	p*K*d = 9.80
	Seed 2	p*K*d = 7.76	p*K*d = 7.41	p*K*d = 7.64
	Seed 3	p*K*d = 8.68	p*K*d = 7.79	p*K*d = 8.00

Figure 1B: Clomipramine template molecule, with its derived seeds and the top 3 *de novo* designed structures having the highest LBA (pKd) for each respective seed

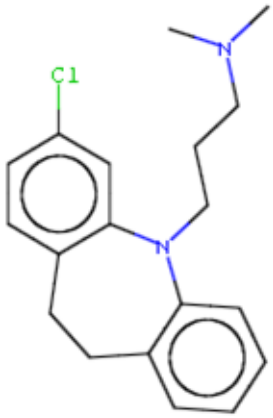
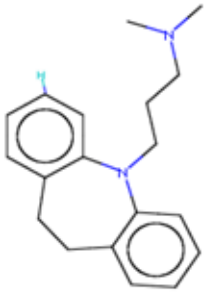
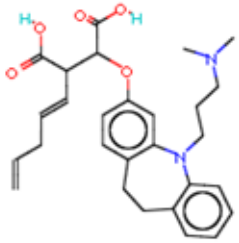
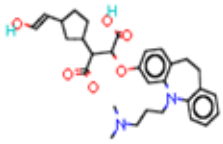
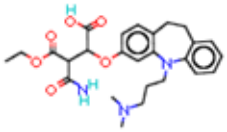
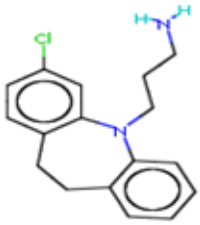
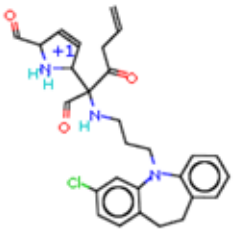
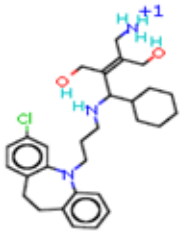
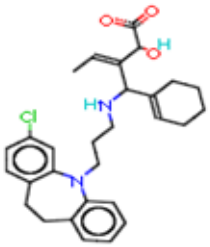
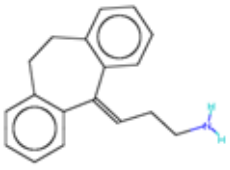
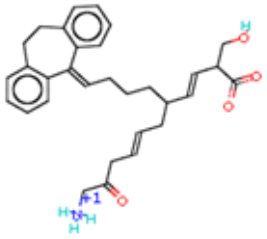
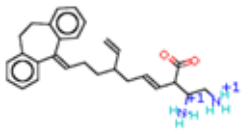
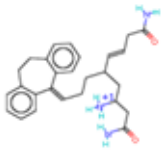
Clomipramine		Novel *de novo* designed structures with highest LBA (p*K*d)		
		Family 1	Family 2	Family 3
 p*K*d = 6.23		 p*K*d = 7.22	 p*K*d = 7.08	 p*K*d = 6.62
	Seed 2 	 p*K*d = 8.30	 p*K*d = 8.22	 p*K*d = 8.05

Figure 1C: Amitriptyline template molecule, with its derived seeds and the top 3 *de novo* designed structures having the highest LBA (pKd) for each respective seed

Amitriptyline		Novel *de novo* designed structures with highest LBA (p*K*d)		
		Family 1	Family 2	Family 3
	Seed 1 	 p*K*d = 9.36	 p*K*d = 9.03	 p*K*d = 8.11

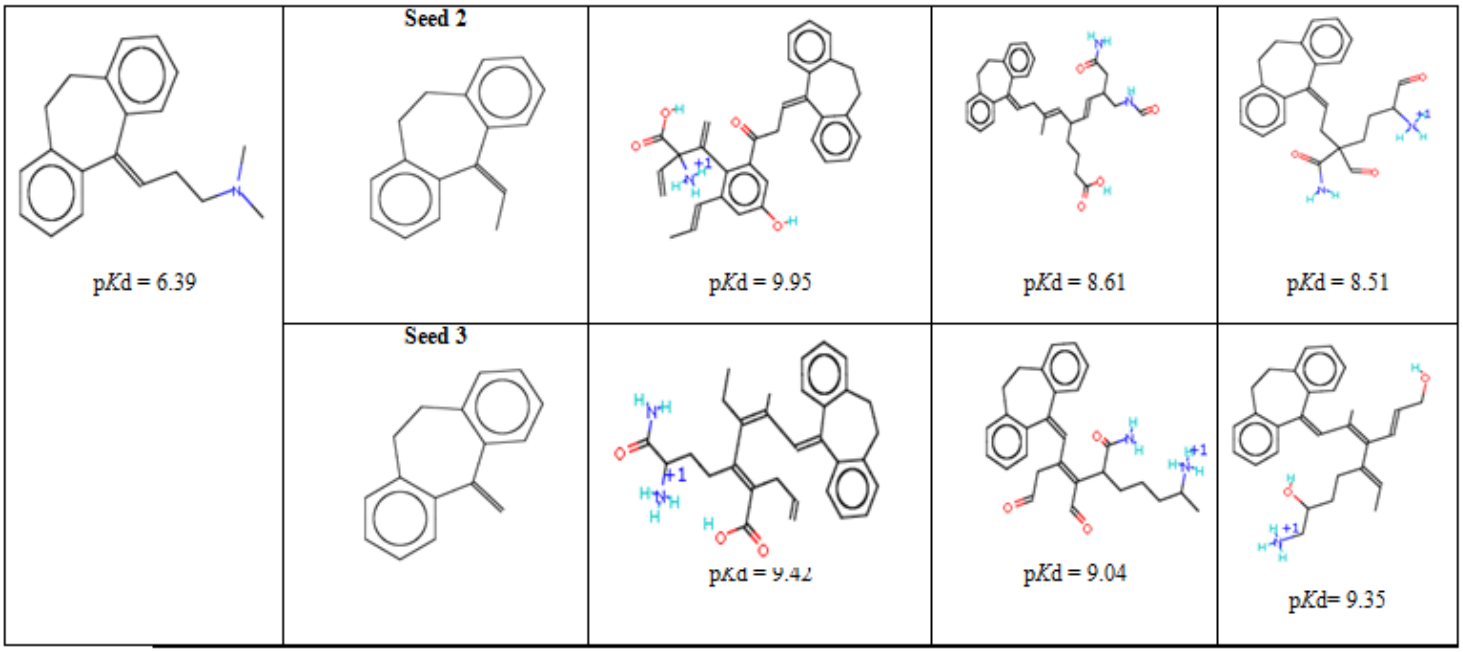


Figure 1D: Fluoxetine template molecule, with its derived seeds and the top 3 *de novo* designed structures having the highest LBA (pKd) for each respective seed

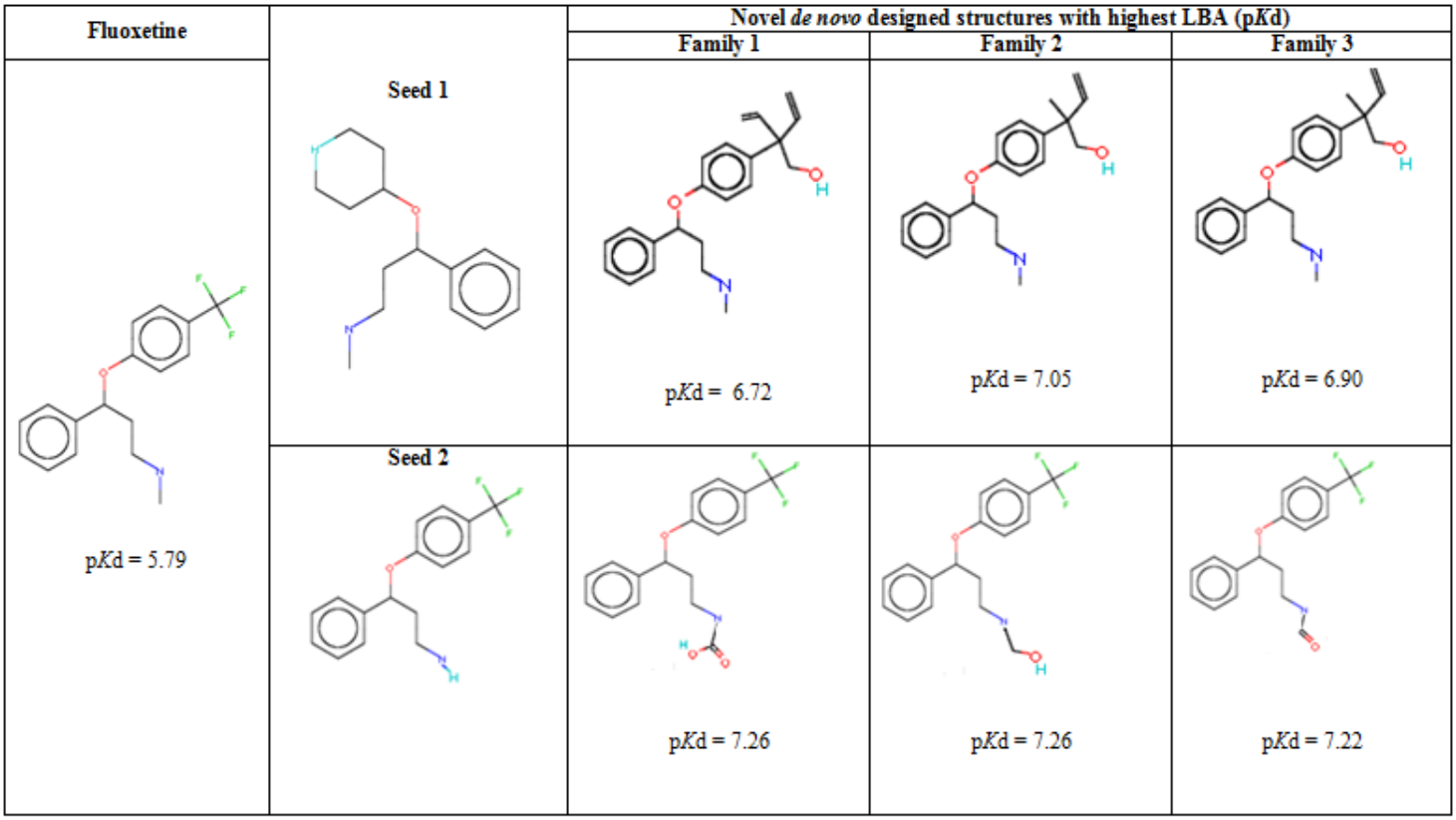
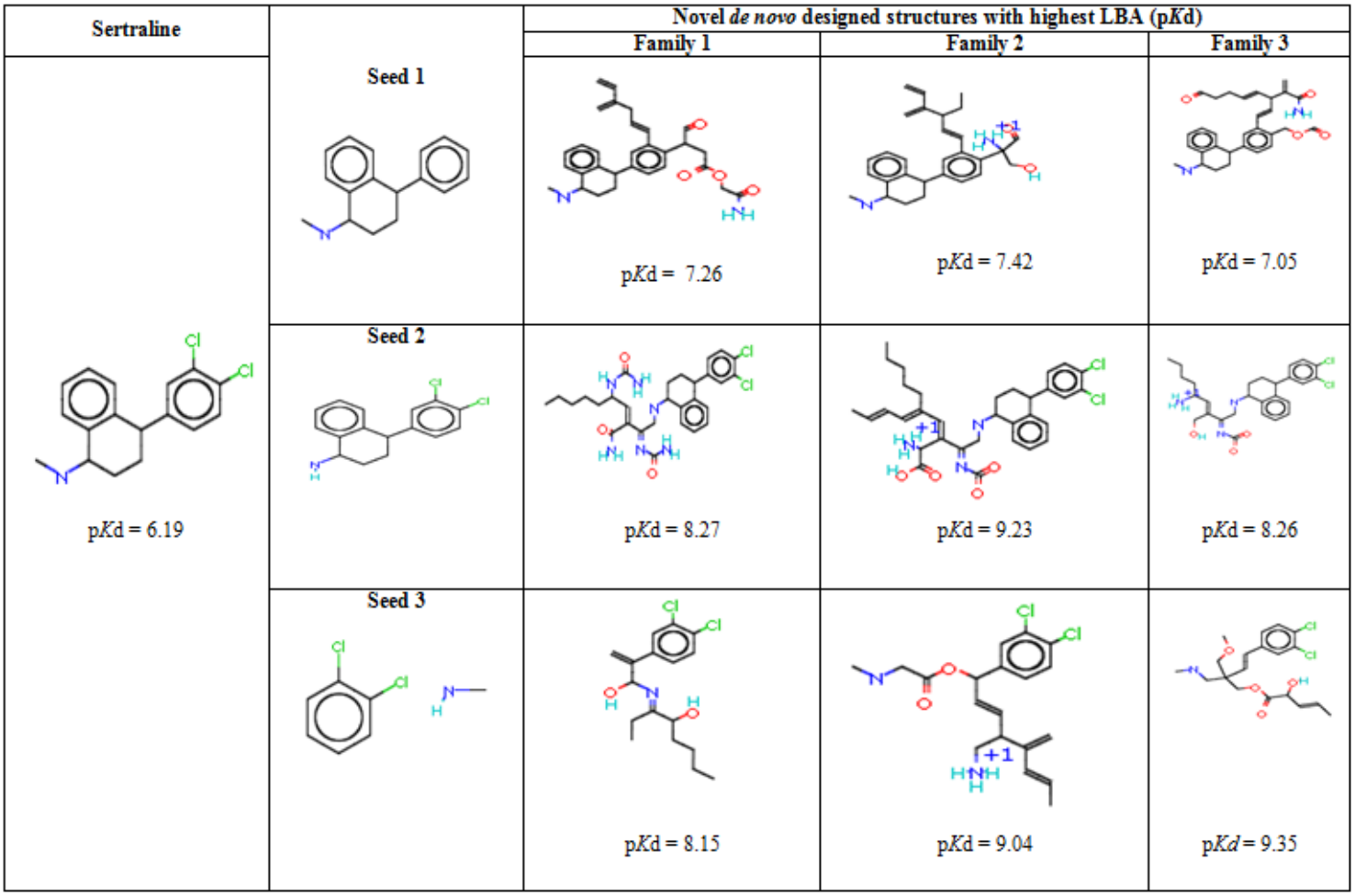


Figure 1E: Sertraline template molecule, with its derived seeds and the top 3 *de novo* designed structures having the highest LBA (pKd) for each respective seed



Discussion:

The results obtained from this study were analysed against a scenario of a perceived necessity for a reduction in novel clinically viable drug turnover time. Consequently, the properties of the molecular cohorts obtained through the parallel implementation of bioisosteric modification and *de novo in silico* drug design were compared with an emphasis on LBA (p*K*d) and Lipinski Rule[16] compliance.

The SSRI and TCA template molecules selected for this study were identified from the PDB bound to different transport proteins. The nature of the bound transport protein was not considered to be critical for this study, owing to the fact that their importance was solely the provision of a rigid pharmacophoric space within which bioisosteric modification of their cognate bound small molecules and *de novo* design of novel structure from constructed seed molecules could be rationally sustained.

Comparison of the LBA (p*K*d) between the parent molecule, and the bioisosterically modified and the *de novo* designed molecular cohort was one of the cornerstones of this study. LBA (p*K*d) calculation of the parent molecules and the bioisosterically modified molecular cohort for their cognate receptor was carried out in X-SCORE® v1.3[14], while that of the *de novo* designed molecular cohort for the same transport proteins was carried out using LigBuilder® v1.2[13]. The fact that both X-SCORE® v1.3[14] and LigBuilder® v1.2[13] were developed by the same workers, and that they share an identical algorithm for LBA (p*K*d) estimation facilitated this comparative exercise.

A number of conclusions may be drawn from the data generated from this study. Analysis of the LBA (p*K*d) data shows that subsequent to bioisosteric molecular modification, no significant increases in LBA (p*k*K) are recorded. Reference is made to Tables 2 and 3 in which, it is evident that the average LBA (p*K*d) for the bioisosterically modified molecular cohort remains similar to that of the parent template ligands (Table 2) and where (Table 3) it may also be seen that merging the highest ranking BIF%[15] moieties from successive modifications also fails to increase the LBA (p*K*d) values.

These results may be explained when the nature of bioisosteric modification is taken into account.

Bioisosteric modification is essentially, the exchange of one, or a group of atoms with others that are similar from an electronic and 3D volume perspective such that new molecules with biological activity similar to the parent are created[17]. The implication of this is that no significantly different atomic interactions will be forged by the bioisosterically modified structures and the receptor when compared to the parent molecule. Consequently, there should be no significant differences in the calculated LBA (p*K*d) between the novel and the parent structure.

When the *de novo* approach was adopted, LBA (p*K*d) data differed markedly from that of the parent molecules, with significant increases being observed in all cases. Reference is made to Figures 1A – 1E in which the *in silico* calculated LBA (p*K*d) for the highest ranking ligands from the first 3 molecular families derived from each template molecule are compared.

These results may also be explained when the modality of the *de novo* approach is considered. Here, the molecular moieties considered critical for binding were planted within the receptor LBP with novel molecular growth being allowed in non critical *loci*. This molecular growth was designed to ensure optimal interaction within the LBP, and also to completely occupy available space in the simplest way possible such that the number of interactions forged between the small molecule and the amino acids forming the LBP perimeter would be maximized without compromising Lipinski's Rules[16] from a molecular weight perspective[18]. This approach consequently allowed both for the identification of high affinity ligands, and also molecular innovation in a way that the bioisosteric approach, which is limited in the number of molecular moieties that could replace targeted *loci* (from an electronic and 3D volume perspective), could not sustain. In fact, a comparison of the molecular cohorts obtained through both approaches (refer to Table 3 and Figure 1A-1E) is indicative of the wider pharmacophoric space explored through the *de novo* approach.

This study therefore practically illustrates and compares, the different thrust of the two drug design approaches both of which continue to play an important role in contemporary rational drug design processes. Bioisosteric replacement is a faster approach that does not contribute significantly to producing higher affinity ligands when compared to a lead molecule and consequently should not be the design modality of choice if a biological scenario of competitive inhibition is being envisaged. Neither should it be used if total innovation is being sought. It is however, relevant if small molecular changes are desired to acquire new patents or to achieve more favourable toxicity profiles, biological activity, or pharmacokinetics. The *de novo* approach, as demonstrated in this study, sustains both higher affinity with respect to a template lead structure and also innovation from a structural perspective. The implication is that it is relevant in the context of creating competitive inhibitors and also in the investigation of hitherto unexplored pharmacophoric space. The disadvantage from a drug design perspective could be an increased discovery turnover time owing to the increased requirement for clinical, toxicity and safety assessment.

References:

1) Twyman R. Rational drug design [Internet]. UK: Wellcome Trust: The Human Genome; c2002. [cited 2014 Jul 19] Available from: http://genome.wellcome.ac.uk/doc_wtd020912.html
2) Huang HJ, Chen CY, Chen HY, Hsu CH, Lee KJ, Tsai FJ. Current developments of computer aided drug design. Journal of Taiwan Institute of Chemical Engineers. 2010; 41: 623-635.
3) Kennewell EA, Willett P, Ducrot P, Luttmann C. Identification of target-specific bioisosteric fragments from ligand-protein crystallographic data. Journal of Computer-Aided Molecular Design. 2006 Jul; 20: 385-394.
4) Ayrton A, Morgan P. Role of transport proteins in drug discovery and development: a pharmaceutical perspective. Xenobiotica. 2008; 38: 676-708.
5) Berman HM, Westbrook J, Feng Z, Gillilang G, Bhat TN, Weissig H et al. The protein data bank. Nucleic Acids Research, 2000; 28: 235-242.
6) Zhou Z, Zhen J, Karpowich N, Law C, Reith M, Wang D. Antidepressant specificity of serotonin transporter suggested by three LeuT--SSRI structures. Nature structural & molecular biology. 2009; 16: 652 - 657.
7) Singh S, Yamashita A, Gouaux E. Antidepressant binding site in a bacterial homologue of neurotransmitter transporters. Nature. 2007; 448: 952 - 956.
8) Nishi K, Curry S, Fukunaga N, Maruyama T, Nakamura T, Nishi K. et al. Structural insights into differences in drug-binding selectivity between two forms of human alpha1-acid glycoprotein genetic variants, the A and F1*S forms. The Journal of Biological Chemistry. 2011; 286: 14427-14434.
9) Güner OF. Pharmacophore perception, development, and use in drug design. Vol 2. California: International University Line; 2000
10) Nadendla RR. Principles of Organic Medicinal Chemistry. New Age International (P) Limited. 2005: 92-94.
11) SYBYL-X 1.2, Tripos International, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA
12) SparkV10, Cresset BioMolecular Discovery Ltd, Broadwater Rd, Welwyn Garden City, UK
13) Wang R, Gao Y, Lai L. LigBuilder: A multi-purpose program for structure-based drug design. J. Mol. Model 2000; 6: 498-516.
14) Wang R, Liu L, Lai L, Tang Y. SCORE: A New Empirical Method for Estimating the Binding Affinity of a Protein-Ligand Complex. J. Mol. Model. 1998; 4: 379-394.
15) Cresset-group. SparkV10. CressetproV10 next generation chemistry software. 2012; 1-88.
16) Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. Adv Drug Deliver Rev. 1997; 23: 3-25.
17) Brown N. Bioisosteres in Medicinal Chemistry. 1st ed. Germany: Wiley-VCH Verlag & Co; 2012.
18) Pegg S, Haresco J, Kuntz I. A genetic algorithm for structure-based de novo design. Journal of computer-aided molecular design. 2001; 15: 911 - 933.