Organic & Biomolecular Chemistry

Cite this: *Org. Biomol. Chem.*, 2011, **9**, 1169

www.rsc.org/obc

PAPER

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†

Yushma Bhurruth-Alcor,[a] Therese Røst,[b] Michael R. Jorgensen,[a] Christos Kontogiorgis,[a] Jon Skorve,[b] Robert G. Cooper,[d] Joseph M. Sheridan,[d] William D. O. Hamilton,[d] Jonathan R. Heal,[d] Rolf K. Berge[b,c] and Andrew D. Miller*[a]

Received 19th May 2010, Accepted 7th October 2010
DOI: 10.1039/c0ob00146e

Peroxisome proliferator activated receptors (PPARs) have been shown to have critical roles in fatty acid oxidation, triglyceride synthesis, and lipid metabolism - making them an important target in drug discovery. Here we describe the *in silico* design, synthesis and *in vitro* characterisation of a novel series of 2,5-disubstituted indoles as PPARα/γ dual agonists. PPAR activation assays are performed with known agonists diazabenzene (WY14.643), aminopyridine (BRL49653) and bisaryl (L165.041), as positive controls. All the indole compounds synthesized are found to be active PPARα and PPARγ agonists, with particular efficacy from those with 2-naphthylmethyl substitution. This is a useful demonstration of a new *de novo* design methodology implemented by the PROTOBUILD program and its ability to rapidly produce novel modulators for a well characterized drug target.

Introduction

Metabolic syndrome is a complex set of disorders encompassing: obesity, dyslipidaemia, insulin resistance, blood pressure elevation and coronary heart disease.[1,2] People with metabolic syndrome have a five-fold greater risk of developing diabetes type II.[3] Peroxisome proliferator-activated receptors (PPARs) have been established as a primary drug discovery target for the treatment of metabolic diseases because of their critical role in the regulation of lipid metabolism and fat cell differentiation.[4–8] These nuclear receptors are transcription factors and bind to specific peroxisome proliferator response elements (PPREs) within promoters.[9–11] The three identified subtypes (PPARα, PPARδ and PPARγ) are all key regulators of lipid metabolism and act by controlling the expression of several genes involved in peroxisomal and mitochondrial fatty acid oxidation, fatty acid uptake and transport, as well as differentiation of adipocytes.[12–14] The main roles of the different PPARs are summarised in Table 1.[14,15] Natural ligands of the PPARs include a wide variety of saturated or unsaturated fatty acids and eicosanoid derivatives.[17] Synthetic PPAR agonists are used in clinical practice, for example fibrates (PPARα agonists) are used to improve dyslipidemia during the treatment of atherosclerosis.[24] The thiazolidinediones are a well-founded class of anti-diabetic drugs that act as PPARγ agonists, thereby improving insulin sensitivity and decreasing the hepatic glucose output.[25,26] Unfortunately the use of thiazolidinediones is known to have attendant side effects such as weight gain, oedema, and anaemia with possible liver dysfunction.[18,27,28] There have thus been suggestions that dual or even pan PPAR agonists (targeting all three subtypes) should be developed, as these may exert complementary and synergistic actions in improving lipid homeostasis and insulin sensitivity, while hopefully reducing attendant side-effects.[29,30]

Here we describe the computer-aided design of a set of novel indole based compounds as dual PPARa/γ agonists, making use of a new *de novo* fragment-based design software program known as PROTOBUILD. PROTOBUILD software seeks to construct new, candidate ligands appropriate to a selected protein receptor ligand-binding site by searching chemical space within the context of the binding site in order to define suitable pharmacophore structures against which series of candidate ligands can be ranked in order of their predicted ligand-binding site affinity. Although this general concept has been implemented already in several other design programs.[31–34] PROTOBUILD differs by making use of a genetic algorithm (GA) that controls both fragment addition and deletions based on dynamic selection pressures. Such an approach allows for a more thorough search of chemical space in particular

[a]*Imperial College Genetic Therapies Centre, Department of Chemistry, Flowers Building, Armstrong Road, Imperial College London, London, UK, SW7 2AZ. E-mail: a.miller07@btinternet.com; Tel: +44 787 963 5513*
[b]*Institute of Medicine, University of Bergen, N-5021, Bergen, Norway*
[c]*Department of Heart Disease, Haukeland University Hospital, N-5021, Bergen, Norway*
[d]*Prosarix Ltd, Newton Hall Town Street, Newton, Cambridge, UK, CB22 7ZE*

† Electronic supplementary information (ESI) available: Experimental details and X-ray crystallography. CCDC reference numbers 667594 (**22**) and 667595 (**23**). For ESI and crystallographic data in CIF or other electronic format see DOI: 10.1039/c0ob00146e

Table 1 Summary of locations and roles of PPARs α, γ and δ

PPAR	Site of expression[16]	Role
α	Catabolically active tissues such as liver, heart, kidneys and skeletal muscle	Oxidation of fatty acids and lipoprotein metabolism[17]
γ	Broad range of metabolically active tissues such as skeletal muscle, kidneys, intestine and adipose tissue	Fat cell differentiation and triglyceride synthesis[15,18–21]
δ	Wide range of tissues and cells, high expression in the brain	Regulation of lipid metabolism and cholesterol efflux[22,23]

by helping to avoid local minima that otherwise develop due to the early inclusion of heavier, hydrophobic fragments at the expense of smaller fragments. The PROTOBUILD approach is an obvious alternative to the Metropolis criterion for Monte Carlo that is implemented in other *de novo* design programs.[31,32] Moreover the use of the GA provides a convenient route for searching different ligand-binding sites with variable affinities, owing to the fact that each binding site can compete for emerging ligand populations *via* the GA. PROTOBUILD also contains several other features that differentiate it from other previously reported design programs. Specifically, the program was developed to exploit a new scoring function, PROTOSCORE, that includes improved terms for the estimation of entropy plus terms for various non-bonded interactions not otherwise represented in design programs. Moreover, PROTOBUILD was designed to take into account protein receptor binding site flexibilities thereby providing an even more realistic view of protein receptor-ligand molecular recognition and binding events. Yet other features involve:

• The use of detailed fragment connection data extracted from known compounds in order to control how the same fragments are connected together within PROTOBUILD. This approach was introduced in order to ensure that new, candidate ligands should be readily synthesized when required.

• The use of pharmacophore constraints and/or the SMILES or SMARTS pattern language in order to direct the generation of new ligands.

• The use of a novel "virtual ligand" (VL) feature in order to allow a user to embed the binding mode(s) of existing biologically active ligands into the program and then use shape compliance as part of an objective function during the generation of new ligands.

Herein, we describe PROTOBUILD in more detail and explain how this program was employed to suggest a novel set of 2,5-substituted indoles as dual PPARα/γ agonists that were then evaluated in an *in vitro* assay system.

Results and discussion

PROTOBUILD data input and program setup

Initially, the PROTOBUILD operation requires a design template and a number of setup procedures. This functionality is provided by a programming module known as SiteAnalysis, written in the SVL language (support vector language) within MOE (Chemical Computing Group).[35] Initially, SiteAnalysis computes a grid-based representation of a specified protein receptor ligand-binding site from the given protein structure. In this process preferred site points for hydrogen bond acceptors and donors are calculated. Where a flexible ligand-binding site is to be used, SiteAnalysis is able to generate an ensemble of receptor conformations. Briefly, a backbone dependent- rotamer library is used to enumerate all possible low energy combinations of side chains in the designated ligand-binding site. A diverse population of low energy combinations is then subjected to SiteAnalysis routines that serve as inputs to PROTOBUILD. By separating protein receptor flexibility calculations from the ligand generation algorithm, multiple receptor states may thereby be included in the GA that controls the growth of new ligands. "Cross-talk" between populations of new, candidate ligands and ligand-binding sites associated with these multiple receptor states then becomes an important contributor towards the growth of new ligands. Clearly this approach does not provide for an exhaustive solution to protein receptor flexibility but does seek to introduce an exploration of low energy ligand-binding site ensembles in a way that is computationally feasible.

The subsequent growth of new ligands may be constrained by the program in a number of user-defined ways, as mentioned in the Introduction. For instance, geometric constraints may also be added in the form of allowed fragment sets for linking. Required fragment sets are placed within a user-defined volume inside the ligand-binding site of interest and are used preferentially by PROTOBUILD. Accordingly, if ligands are grown to within a predefined distance of the geometrically constrained fragment sets, then a bond is created to include the required fragment sets within growing ligand structures thereafter. Further constraints may be applied by the specification of pharmacophore constraints that include locations of cationic, anionic, hydrophobic, H-bond donor, H-bond acceptor, aromatic and 'any' other feature types of interest. Constraints may also be applied through the execution of SMILES or SMARTS queries of growing ligands that are intended to specify molecular structural elements or templates that must be included or excluded from a final population of ligands grown by the GA. Otherwise if required, a virtual ligand (VL) command constrains ligands generated to a shape(s) defined by an existing ligand binding mode(s). In order to support the VL function within PROTOBUILD, the GA has the ability to use X-ray crystal structure co-ordinates of ligands co-crystallized with protein receptors or else co-ordinates generated by the docking of a ligand of choice into a given ligand-binding site of a protein receptor of interest whose X-ray crystal structure is known. These co-ordinates provide the means to imprint the preferred VL volume of a ligand of choice into the binding site definition. The VL volume is similar to an included pharmacophore volume but the rules for adherence to this representation are programmable and more complex. This allows the user to fine-tune the required compliance for any candidate ligand to this volume and thus its ability to participate in further growth events.

PROTOBUILD ligand generation

In order to generate new ligands, PROTOBUILD makes use of GA-based search and optimization logic for fragment growth under

user-defined constraints (as noted above). Evolutionary search techniques have been shown to be effective at identifying energy minima in a number of complex (NP-hard) problems within computational chemistry[36] and indeed have been used within other *de novo* design program implementations too.[37] Implementation of our GA-based approach requires that initial molecular fragment 'seeds' should be supplied to the program in order to initiate ligand generation. Usually these fragment seeds are extracted from data of whole ligand docking studies or from X-ray crystallography studies of co-crystallized ligands/protein receptors. So too seeds may derive from fragment docking studies. In PROTOBUILD the GA then generates and evolves ligands of increasing fitness, as determined by a predicted binding score. Starting from the fragment seeds, ligands are grown by the steady incorporation of new added fragments from a program-accessible fragment library that link fragment seeds together as ligand structures are generated. The conformational space of new added fragments is explored by systematically evaluating all rotamers at defaulted 5° angles before low energy solutions are identified and retained. New candidate ligands that result are then passed to a module for fitness evaluation. The PROTOBUILD program and associated code libraries are written using ISO/ANSI C++ and compiled using the Visual C++ 8.0 compiler. Command architecture and parameters are defined by XML. Within PROTOBUILD bond lengths, angles and atomic radii are provided by the Tripos 5.5 forcefield.[38] In a structure-based design scenario, final PROTOBUILD designs are minimized using the MMFF94 forcefield.[39] The latter has been shown to perform well elsewhere for similar tasks.[40]

Fitness evaluation is performed using an empirical scoring function PROTOSCORE (see below) that ranks the effectiveness of receptor-ligand binding interactions for each new ligand conformers generated by program under GA guidance. A variety of penalties are then used to filter the ligand list prior to additional rounds of ligand growth. As noted above, the penalty criteria can include absence of pharmacophore features, presence of chiral centers, secondary and tertiary branch points, Lipinski parameters (including molecular weight and log P upper limits), number of linked ring systems and so on. The penalty criteria are particularly helpful in removing large numbers of candidate ligands with overcomplicated structures even though they may result previously from valid fragment pairings. Clearly penalty criteria could vary significantly depending upon the protein receptor and ligand-binding sites under investigation. Once the high scoring ligands have been identified for the next round of ligand growth, a subset of the remainder is still recycled in the next round allowing some of the less desirable ligands a "wildcard entry" to the next growth and search cycle. In particular, candidate ligands are retained with a low degree of structural similarity in comparison to compounds with the highest score from PROTOSCORE. In so doing, diversity is increased in acknowledgement of the fact that some ligands whilst being suboptimal may in fact become much better ligands in subsequent growth and search cycles after further fragment additions.

An inherent problem during *de novo* design is the inclination to generate compounds with excessive hydrophobicity.[32] Such compounds are undesirable for a number of reasons such as poor solubility and hence drug formulation. This problem arises within ordinary growth algorithms due to the tendency of the scoring function to favor the addition of large and usually hydrophobic fragments to the growing molecule. This is because these compounds score comparatively well due to the Van der Waal's interactions with the receptor and thus are retained for further growth. However, such ligands are undesirable in terms of physical properties and seldom represent a global minimum. Fortunately, the extensive use of penalty criteria such as Log P selection helps to overcome this problem. Also, added fragments can be associated with a user-defined selection pressure that may be used to control the likelihood of selection at any given time during ligand growth. In particular, PROTOBUILD is able to alter dynamically the selection pressures of the added fragments thus preventing the "flood-filling" of the ligand-binding site early on in the design process with hydrophobic fragments. There is also a facility to prevent or limit the number of ring systems that are directly connected to each other because these systems often present challenging synthetic routes. Alternatively, SMARTS patterns may be used to explore only ligands of certain generalized scaffolds and force solutions that adhere to these rules.

PROTOBUILD can run in two different modes. In one 'deterministic' mode a set number of iterations of the GA are executed. In the other 'non-deterministic' mode, the run continues indefinitely. In both modes the program may be paused such that emerging ligands may be reviewed and run parameters may be changed in accordance with a runtime programming language implemented in XML. Both the deterministic and non-deterministic search cycle modes for PROTOBUILD can make use of a fragmentation capability in order to boost the exploration of diversity in chemical space exploration during a given PROTOBUILD run. In fact, even entire ligands may be supplied to PROTOBUILD as seeds where they are broken up randomly and recursively in order to generate a controllable number of discrete starting points for the design process. This strategy is particularly useful in lead optimization studies.

Scoring with PROTOSCORE

Central to PROTOBUILD is the empirical scoring function that was newly developed for use in PROTOBUILD. This scoring function is known as PROTOSCORE. PROTOSCORE incorporates a treatment for ligand solvation, receptor deformation upon binding and calculations for aromatic and heterocyclic ring interactions. PROTOSCORE builds upon many of the concepts implemented in the earlier pioneering work of Bohm.[41] The main equation developed for PROTOSCORE is shown (1):

$$\begin{aligned} p\mathit{K}_i = c + {}_{x1}\mathrm{HB} + {}_{x2}\mathrm{SURFMATCH} + {}_{x4}\mathrm{LIGFLEX} \\ + {}_{x5}\mathrm{PROTFLEX} + {}_{x6}\mathrm{SOLV} + {}_{x7}\mathrm{ARO} \end{aligned} \quad (1)$$

The terms in the equation will be described briefly. Hydrogen bonds (HB) are calculated according to the term (2):

$$\mathrm{HB} = \Sigma\,\mathrm{hbF} + \Sigma\,\mathrm{hbwF} + \Sigma\,\mathrm{hbmF} \quad (2)$$

where hb is the interaction energy of a perfect hydrogen bond and F is a measure of geometric deviation from ideal with a value between 0 and 1. The terms hbw and hbm are interaction energies for water mediated hydrogen bonds and transition metal contacts respectively. Surface lipophilic interactions and polar/apolar interactions (SURFMATCH) are calculated by means of a surface-matching algorithm described by the following eqn (3):

$$\mathrm{SURFMATCH} = \Sigma(\text{lipo–lipo}) - \Sigma(\text{alipo-lipo}) - \Sigma(\text{lipo-alipo}) \quad (3)$$

where lipo-lipo is the lipophilic contact area between ligand-binding site and ligand, alipo-lipo is the contact area between polar areas of a ligand-binding site and non-polar ligand regions while lipo-alipo is the contact area between non-polar regions of a ligand-binding site and polar ligand regions. SURFMATCH expresses the protein receptor-ligand contact surface area using a grid method in a method similar to that previously reported.[41] Where ligand polar groups are buried in lipophilic environments and ligand-binding site polar groups map close to ligand lipophilic groups, then there is a penalty.

Terms LIGFLEX and PROTFLEX are included to account for protein receptor and ligand flexibility. The entropic effect caused by the requirement to reduce conformational degrees of freedom upon binding can be expressed by a count of rotatable bonds for both interacting ligand and protein receptor. For the ligand the term LIGFLEX calculates the number of acyclic sp3-sp3 or sp3-sp2 bonds but does not include terminal groups. Schemes for protein receptors are more complicated because flexibility of the ligand-free state must be calculated and compared with flexibility in the ligand-bound state. After investigation of a number of schemes PROTFLEX calculates the number of rotatable bonds of accessible portions of side chains that contact the binding ligand. A more accurate version of the function PROTFLEX2 pre-calculates the number of low energy rotamers available to ligand-binding site amino acid residue side chains and uses this value to weight the number of rotatable bonds in contacting side chains. This extra step needs only to be run once per protein receptor. Otherwise, SOLV takes into account the entropic gain due to desolvation when ligand binds the ligand-binding site of a protein receptor. The term SOLV uses a finite difference method for fast solution implementation of the Poisson-Boltzman equation. Only the desolvation volume of bound ligand is calculated. This term is computationally expensive requiring a Poisson-Boltzman equation to be solved twice per ligand. With the current implementation this adds 0.5 s per ligand to the calculation. Finally, aromatic interactions are described by the following equation:

$$\mathrm{ARO} = \mathrm{AROaro} + \mathrm{AROcat} + \mathrm{AROs} \quad (4)$$

Principally, the ARO term (4) contains calculations for aromatic-aromatic rings AROaro, cation-pi interactions AROcat and sulfur-pi interaction AROs. AROaro is calculated by measuring distance and angle between the rings of a ligand and those of Trp, Phe, Tyr or His amino acid residues belonging to a protein receptor. The function calculates energetically favourable edge to face and offset face to face interactions as previously described.[42,43] AROcat calculates cation-pi interactions expressed as a function of the distance and angle of a positive charge to the face of an aromatic ring system. The value of the term is provided from a lookup table generated from a set of experimentally determined interactions.[44] The AROcat term compensates for the penalization of polar-apolar interactions identified within the SURFMATCH term. The separate term AROs describes the interaction phenomena observed when an sp3 sulfur atom interacts with a pi system. Once again the value of this term is trained from experimental data.[45] This interaction can be underestimated using a simple treatment of lipohilic contacts.

Selection of training and test sets

The definition of suitable training and test sets of protein receptor-ligand complexes is as important as the definition of the descriptors to be used during training. For PROTOSCORE a diverse set of 275 receptor-ligand structures was used. This set incorporated kinases, proteases, phosphatases, nuclear receptors, metalloproteases and sugar binding proteins. For this dataset the equation fitted by PLS yielded an $r^2 = 0.82$, $q^2 = 0.77$, $S = 6.6$ kJ mol^{-1}, $S_{\mathrm{press}} = 7.4$ kJ mol^{-1}. On a separate test set of 50 proteins an r^2 of 0.73 and S_{press} of 7.9 kJ mol^{-1} was recorded.

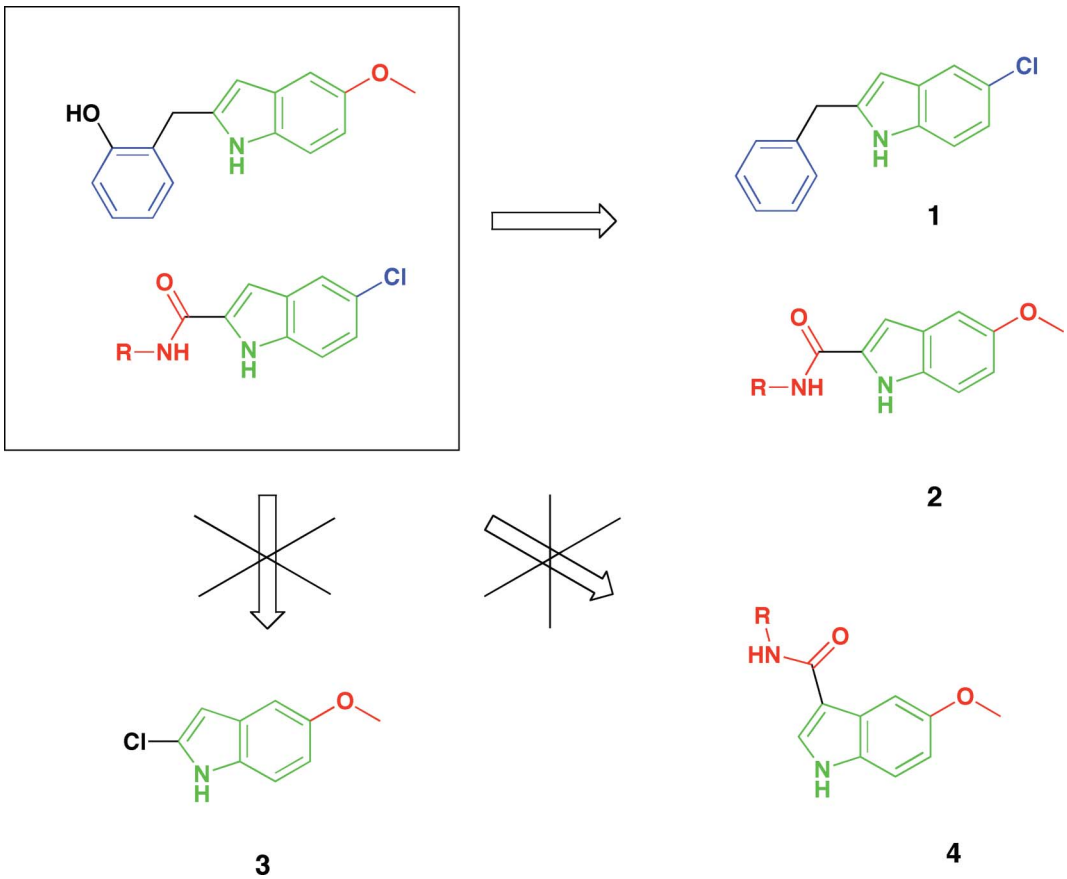
Defining fragments for PROTOBUILD

Ideally a *de novo* design program should only really output candidate ligands with structures that can be realistically synthesized from available reagents. This is not always the case since proposed ligands are constructed from randomly assembled fragments rather than synthons connected by synthetic rules. Some *de novo* design approaches to new ligands have been reported to construct ligands from reagents although it is difficult to express sufficient chemistry *in silico* to provide the level of diversity required in many design projects.[37] Automated retrosynthetic approaches have also been used to assess ease of synthesis but often lead to solutions that are not automatically favoured and which and are often unreliable when dealing with complex substitution patterns on aromatic groups.

PROTOBUILD generates new ligands according to a set of fragment-fragment interconnectivity rules. These rules stipulate that fragment additions can only take place if the fragment-fragment substructure exists in a known compound. The fragment library is constructed by decomposition of 38 000 known, patent protected, drug-like molecules extracted from the Thomson Pharma database. In this analysis the observed substitution patterns for the fragments are stored as are the frequencies of substitutions. Enforcing these rules during fragment coupling ensures that any two combined fragments must be present as a substructure within an existing molecule in the training set (Scheme 1). Any library of molecules can be used as input to the fragment definition procedures. The inclusion of these rules does limit the chemical space being searched but significantly enriches the chemical 'sense' of the resulting molecules. The standard fragment library in PROTOBUILD is comprised of 594 fragments, most of which are ring systems. Within this set of fragments there are 1843 unique atom environments that can yield a theoretical maximum of 3 396 649 different ligands composed of only 2 fragments. Analysis shows however that only 53 426 of these potential ligands can be generated based on the trained fragment pairing rules since the majority of the atom environment pairings are not observed in the training set.

Designing novel indole compounds

At the start of the design process the crystal structure of a known potent PPARγ agonist **1** (GI262570) complexed with PPARγ (PDB 1FM9) was studied (Fig. 1).[46] The crystal structure demonstrates a number of important interactions between the ligand and receptor including hydrogen bonding between the carboxylic acid moiety and residues His449, Tyr473 and His323 (of the AF-2 helix). The hydrophobic phenoxy moiety interacts



Scheme 1 Derivation of fragments from known molecules and rules for generating new ligands. Training compounds shown in the box with different fragments colored. Outputs 1 and 2 are examples of allowed outputs that have the same substitution patterns on the indole as observed in training compounds and fragment A-B combinations are maintained. Outputs 3 and 4 are examples of disallowed products. Output 3 is disallowed due to incorrect fragment pairing whereas output 4 is disallowed due to an incorrect substitution pattern.

Fig. 1 Structure of **1** (GI262570).

primarily through van der Waals forces with residues Tyr327, Leu330 and Cys285. Analysis of this ligand-binding domain (LBD) and also that of PPARα (PDB 1I7G) enabled a hypothesis for a pharmacophore required to confer both PPARγ and PPARα agonism that could be used as an input to a *de novo* design scheme (Fig. 2). The coordinates of the carboxyl group from ligand **1** were used to represent the 'seed' structure for the ligand-growing algorithm.

Two pharmacophore points were specified as being satisfied by output structures, an aromatic feature in the narrow hydrophobic channel located 4.1 Å from the carbon atom of the carboxyl seed and an included volume feature located 10.6 Å from the same carbon atom and 7.1 Å from the aromatic feature. The flexibility of the LBD was assessed by evaluating sidechain rotamer flexibility for all residues with at least one contact to ligand **1** in the PPARγ structure using procedures implemented in MOE (Chemical Computing Group Inc., Quebec, Canada). In general, side chain mobility was found to be much lower in the narrow channel of the binding site proximal to the key hydrogen bonding interactions of the carboxyl group, than the distal regions in the opening of the LBD. This is in agreement with published B-factor values for the crystal structure. The lowest energy rotamer states as determined by AMBER89 forcefield were clustered and 144 were used as inputs into the PROTOBUILD program as competing receptor structures for evolving ligands.

Finally, PROTOBUILD was used to grow molecules in the binding site representations driven by the PROTOSCORE scoring algorithm as the objective function, in addition to other target filters including molecular weight limit <500, predicted solubility $>1\ \mu M$ and number of hydrogen bond donors and acceptors <10.[47,48] A total of 594 distinct fragments were utilised in the growing procedures, each with synthetic *meta*-data providing instructions as to which fragments may combine at which substitution points, in a manner similar to the RECAP scheme.[49] As shown in Fig. 2, the input to the program was a carboxylic acid seed originally excised from the X-ray structure of GI262570 and which was marked as protected such that the algorithm would not delete these atoms during rounds of optimisation. PROTOBUILD was run

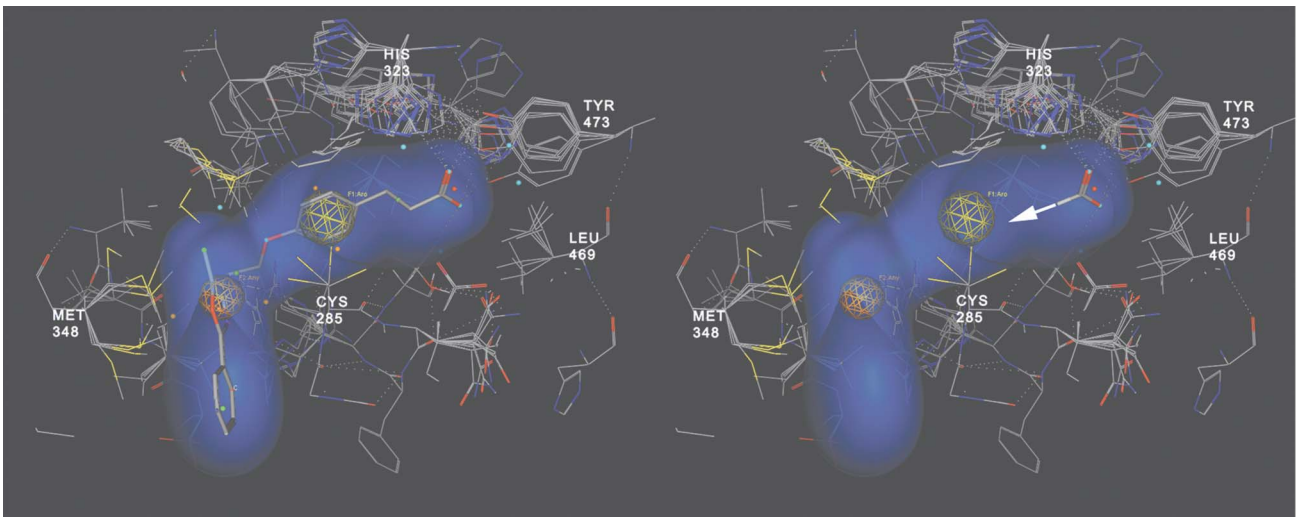


Fig. 2 Design process. **Left**: Ligand **1** (stick format) in complex with PPARγ with overlaid rotamer conformations from receptor flexibility analysis (pharmacophore features shown as wire spheres). **Right**: carboxylic acid seed for *de novo* design shown with exit vector but with ligand **1** removed from site.

under a non-deterministic mode in which iterative populations of ligands were built and trimmed until the algorithm was manually terminated once sufficient numbers of ligands were generated which had met the output criteria. The process resulted in several thousand structures that satisfied the pharmacophore scheme and had predicted K_d binding affinities <1 μM according to the PROTOSCORE function. The solutions were clustered using MACCS key fingerprints and 75% similarity as the cluster definition. This resulted in ~200 different clusters or chemotypes to be assessed for tractability and potential novelty. These results contained known active PPARγ chemotypes as well novel structures. Some of the known PPAR modulator scaffolds observed in the PROTOBUILD results are described (Fig. 3). The observance of known chemotypes encourages that other compounds proposed by the algorithm may also prove to be active. In total three different novel scaffolds were shortlisted for synthesis. These were selected based on highest ligand efficiency values rather than overall predicted affinity since generally it is considered that this often produces simpler starting structures for initial designs. The 2,5- disubstituted indole scaffold was the first scaffold to be synthesised, primarily based upon availability of reagents and the ease of synthesis, coupled with a degree of novelty.

Recently, indole cores have featured in other known PPAR agonists described in the past few years,[50] including several that are comprised of a 1,5-disubstituted indole core.[51–54] By contrast, PROTOBUILD output predicted that particular 2,5-disubstituted indoles, wherein position-2 substitutions are aromatic in character, should also result in effective PPAR agonists. Furthermore, PROTOBUILD output also predicted an effective substitution of hydrogen for fluorine within saturated side-chain substitutions in position-5. Therefore, we resolved to prepare a selection of 2,5-disubstituted indole variants with different aromatic groups, R^1, in the 2-position and an ethoxypropanoic side chain in 5-position, wherein R^2 is either a methyl- or a trifluoromethyl- group according to the PROTOBUILD output (Fig. 4). Such compounds each have one chiral centre α- to their respective carboxylic acid moieties, but we elected to carry out racemic syntheses in the first instance and these racemates were then evaluated as potential PPAR agonists in appropriate PPAR assays *in vitro*.

Table 2 2,5-disubstituted indole compounds **2** to **7** as shown in outline structure in Fig. 4

Compound	R^1	R^2
2	H	CH_3
3	H	CF_3
4	Benzyl	CH_3
5	Benzyl	CF_3
6	2-Naphthylmethyl	CH_3
7	2-Naphthylmethyl	CF_3

Synthetic Chemistry

The PROTOBUILD software program suggested the synthesis of molecules **2** to **7**, with R^1 = H, phenyl and naphthylmethyl groups as potential PPARα/γ modulators (Table 2). Indole structures **2** and **3** were introduced to evaluate the importance of 2-substituents in indoles for binding to the receptor in a hydrophobic site that is occupied by the methyl oxazole moiety of ligand **1**.

For indoles **2** and **3**, where R^1 is hydrogen, commercially available indole-5-carboxaldehyde **8** provided an ideal template from which the aryl substituents and PPAR-binding carboxylic acid moiety could be synthesised (Scheme 2). A Horner–Wadsworth–Emmons (HWE) coupling was carried out with the required phosphonoacetates **9** and **10** to give the *Z* and *E* isomers **11** to **14**. The reduction of the double bond with magnesium turnings in dry methanol proceeded smoothly, as described by Lohray *et al.*[55] Finally, saponification of the methyl ester with potassium hydroxide in a 1 : 1 mixture of ethanol–water gave the desired targets **2** and **3** in excellent yields.[56] The HWE coupling reagents **9** and **10** were prepared as shown (Scheme 3). When R^2 is a methyl group, triethylphosphite and commercially available 2-chloro-2-ethoxyacetic acid **17** were coupled using Arbuzov chemistry as described by Grell and Machleidt.[57] This reaction proceeded smoothly to give phosphonoacetate **9** in a quantitative yield of

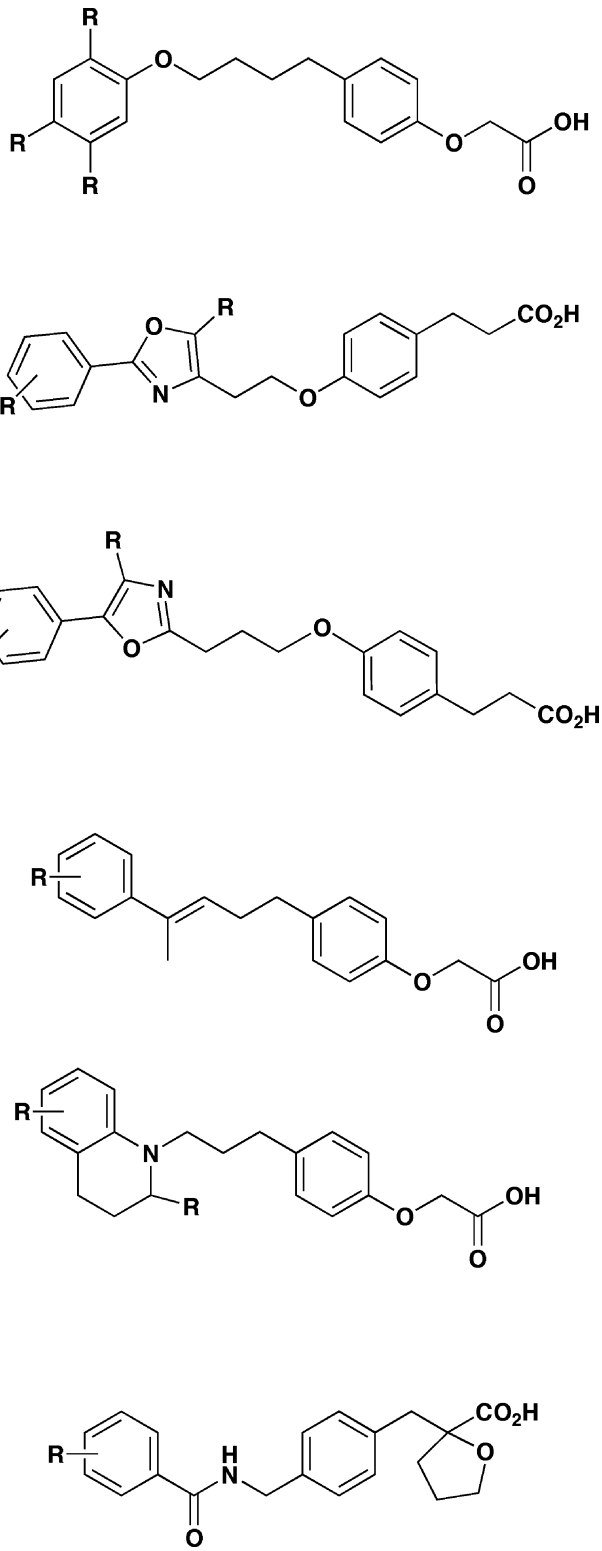


Fig. 3 Some of the previously reported chemotypes identified within PROTOBUILD results.

92% ready for use without any purification. For the synthesis of the 2,2,2-trifluoro phosphonoacetate **10**, tosyl azide was prepared by coupling sodium azide with tosyl chloride.[58] This azide appeared stable to handle and was used to synthesize a diazo intermediate **18** from phosphonoacetate **19** using sodium hydride in dry THF as

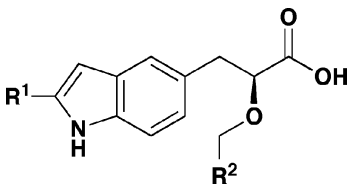


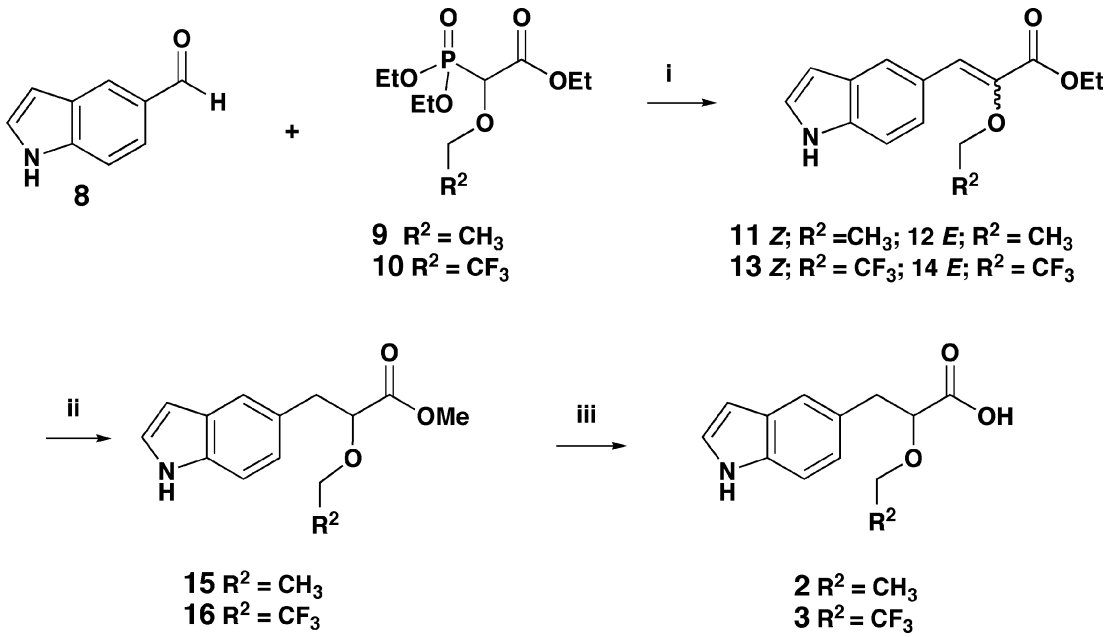
Fig. 4 2,5-Disubstituted indole scaffold where R^1 represents the volume feature, R^2 the aromatic feature and the carboxylate functional group represents the seed.

described by Moody *et al.*[59] and Regitz *at al.*[60] Diazo intermediate **18** proved stable to silica chromatography. Thereafter, rhodium(II) acetate dimer was used to generate a carbenoid from diazo compound **18** which was inserted into the O–H bond of 2,2,2-trifluoroethanol, as per Haigh *et al.* (Scheme 3).[61]

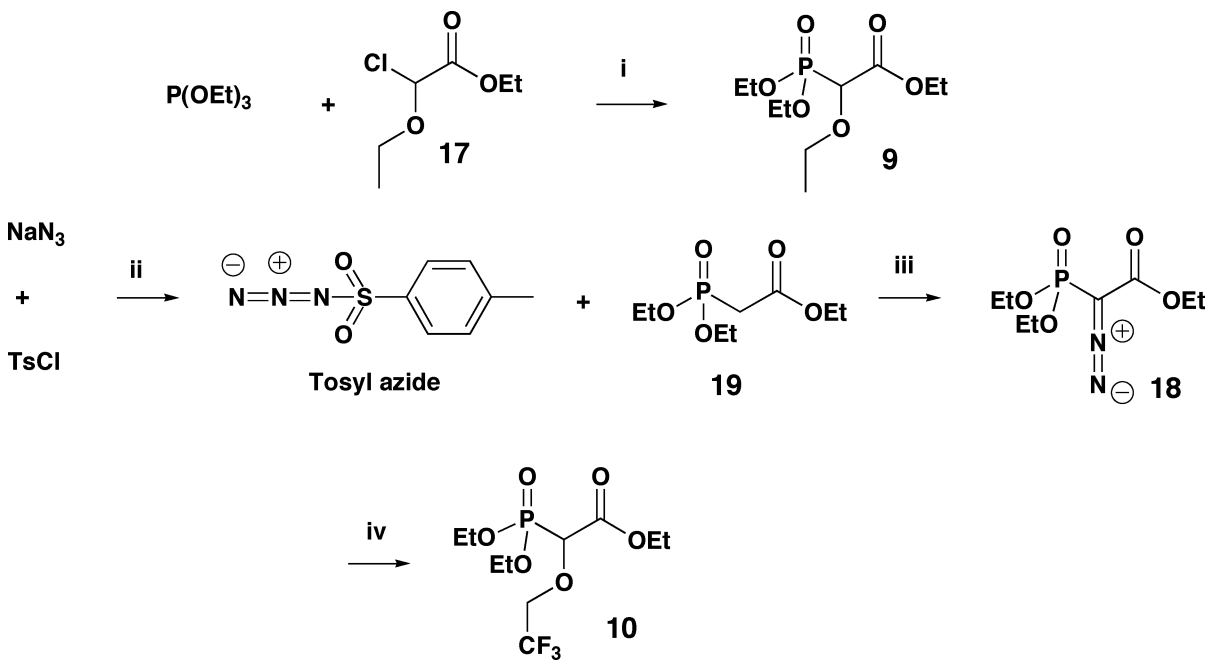
For indoles **4** to **7**, where R^1 is a benzyl or a naphthylmethyl group, a suitably functionalised cyano indole **20** provided an ideal starting material (Scheme 4). The first step was to protect the nitrogen of commercially available indole **20** with a benzene sulfonyl group with the help of the phase transfer catalyst $(n\text{Bu})_4\text{Br}$. Alternatively, benzene sulfonylchloride in THF was found to protect indole **20** with the help of sodium hydride. The former transformation gave the higher yields in our case. Protected indole **21** was purified by crystallisation from ethanol in a 96% yield. The benzene sulfonyl group was chosen as it allows chelation of lithium to the oxygen of the benzene sulfonyl group and enhances substitution on the C2 position due to the influence of the nitrogen atom. After treatment of indole **21** with LDA in dry THF at −78 °C adapted from a method by Sundberg and Russell, the lithiated intermediate (without isolation) was treated with either benzyl bromide or 2-(bromomethyl)-naphthalene.[62] After chromatography on silica gel, crystallisation from hexane–DCM 20 : 1 yielded shiny white crystals. X-ray crystal structure analyses were carried out to confirm the selective introduction of the R^1 group on the 2-position of the indole (Fig. 5). The next step required deprotection of the benzene sulfonyl group in basic conditions. Cyano indoles **24** and **25** were subsequently reduced to the corresponding aldehydes **26** and **27** respectively using DIBAL. Next, a series of HWE couplings were performed to couple aldehydes **26** or **27** with phosphonoacetates **9** or **10** resulting in four different *Z* and four different *E* double bond isomers **28–35**. Double bond reduction with magnesium turnings in dry methanol was then performed giving four indole precursors **36** to **39** that were saponified as described previously to yield the desired indole compounds **4** to **7**. Hence the six desired indoles **2** to **7** were thus successfully synthesised. The purity of the compounds was an important aspect, prior to their testing *in vitro*. A minimum of 97% purity of the indoles was achieved by careful purification of all the intermediates, followed by crystallisation wherever possible.

In vitro PPAR activation assays

The activation of the three PPAR subtypes by the indoles **2** to **7** was studied *in vitro* using a luciferase reporter gene assay in which the reporter construct contained PPREs in the promoter of the firefly luciferase gene. Specifically, a plasmid encoding for each PPAR subtype, as well as the luciferase reporter gene was transfected into human breast cancer MCF-7 cells. Transfection was performed in the absence of PPAR expression vectors as negative control.

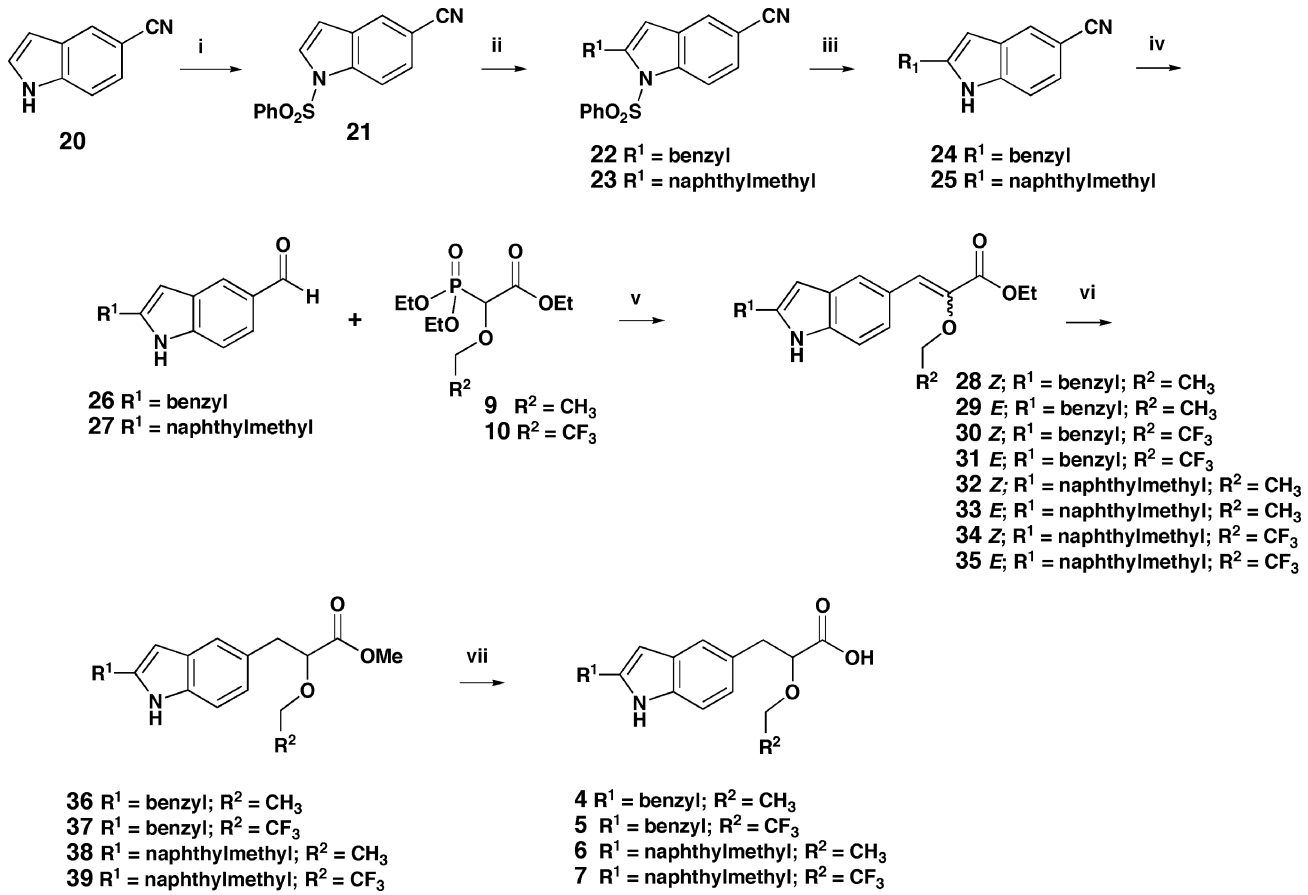


Scheme 2 Synthesis of indoles **2** and **3**. i) NaH, dry THF, (0 °C to) ambient temperature, 26–45 h, 93–99% ii) Mg turnings, dry MeOH, ambient temperature, 3 h, 60–83% iii) KOH, EtOH–H_2O 1 : 1, reflux, 15 h, 92–93%.



Scheme 3 Preparation of the phosphonoacetates **9** and **10**, i) Anhydrous DMF, reflux at 145 °C, 15 h, 92% ii) EtOH, acetone, ambient temperature, 23 h, 98% iii) NaH, dry THF, 0 °C, 16 h, 73% iv) 2,2,2-trifluoroethanol, $[Rh(OAc)_2]_2$, benzene, reflux, 22 h, 63%.

This allowed checking for any endogenous PPAR activity. Under these conditions, no significant luciferase expression was detected. A negative control was included for each set of experiments by treating cells with dimethyl sulfoxide (DMSO) in the absence of any compounds. Positive controls were performed with the following selective potent PPAR agonists at concentrations of **40** (WY14.643) at 30 μM for PPARα; **41** (BRL49653-rosiglitazone) at 1 μM for PPARγ, and **42** (L165.041) at 1 μM for PPARδ (Fig. 6). Experiments were conducted with five different concentrations of each indole **2–7** at 1 μM to 75 μM in DMSO to obtain a preliminary indication of their effects on PPAR activation. The lowest concentration that gave activity as well as a higher concentration relative to the first was selected for data presentation (Fig. 7). All reported experiments were performed in triplicates at these indicated concentrations. All results were normalised relative to the control and the standard deviation was calculated from the variability of the readings obtained for one compound at a particular concentration.



Scheme 4 i) $PhSO_2Cl$, $(n\text{-}C_4H_9)_4NBr$, NaOH(aq), toluene, H_2O, ambient temperature, 25 h, 96% ii) a) LDA, −78 °C, dry THF, 55 min: b) arylBr, −78 to 0 °C to ambient temperature, 16 h, 30–66.5% iii) NaOH(aq), MeOH, reflux, 18 h, 99% iv) DIBAL, dry DCM, 0 °C to ambient temperature, 16 h, then 0 °C, 4.5 h, 80–86% v) NaH, dry THF, (0 °C to) ambient temperature, 26–45 h, 78–90% vi) Mg turnings, dry MeOH, ambient temperature, 3 h, 73–80% vii) KOH, EtOH–H_2O, reflux, 15 h, 93–99%.

Structure activity relationship

When comparing the PPAR activation profiles of the 6 synthesised indoles, the following patterns are apparent:

1. Indoles **2** and **3** were not capable of significantly higher PPAR activation than any of the three indicated positive controls **40–42**.

2. Indole **4** significantly activated PPARα relative to positive control **40** (WY14.643), equivalent to **41** (BRL49653-rosiglitazone) with respect to PPARγ-selective activation, but much less significant than **42** (L165.041) with regard to PPARδ–selective activation. By contrast, Indole **5** was equivalent with but did not exhibit significantly higher PPAR activation than any of the three indicated positive controls **40–42**.

3. Indole **6** induced PPARα selective activation more than the positive control **40** (WY14.643), equivalent with **41** (BRL49653-rosiglitazone) with respect to PPARγ-selective activation, but also less significant than **42** (L165.041) with regard to PPARδ–selective activation.

4. Indole **7** also activated PPARα more than positive control **40** (WY14.643), but clearly less effective than **41** (BRL49653-rosiglitazone) and **42** (L165.041) with regard to PPARγ and PPARδ-selective activation respectively.

Hence, our initial conclusions from this limited series of compounds are that inclusion of a 2-naphthylmethyl functional group appears to confer better indole-mediated PPARα activation than the inclusion of the 2-benzyl group. In addition, in comparing 2-naphthylmethyl indoles **4**, **5**, **6** and **7**, the overall functional requirement for a PPARα/γ selective agonist appears better satisfied by indoles **4** and **6** so implicating the ethoxy functional group as a potentially more suitable functional group than 2, 2, 2-trifluoroethoxy. The PROTOBUILD output indicated that the indole ring could be tolerated in the ligand binding domains of both PPARα and PPARγ. The ligand binding domains of PPARα and PPARγ are similar in size, allowing such type of agonists to be developed.[63,64] The predicted binding mode of the indole **6** involves the 2-naphthylmethyl group fitting into the hydrophobic site near the entrance of the LBD pocket; the indole ring fitting into the narrow hydrophobic channel occupied by the phenoxy moiety of ligand **1** while the carboxylic acid occupies the same position as with **1** and other PPAR activators (Fig. 8). The interaction of the 2-naphthylmethyl compounds with receptors was scored highest with interacting sidechain rotamers positioned as shown (Fig. 2), with the exception of the flexible Met348 residue whose optimal rotamer-position occupied a different conformation.

From now, we aim to carry out some further tests on the compounds, including toxicity and some preliminary pharmacokinetic measurements in order to select the compound with

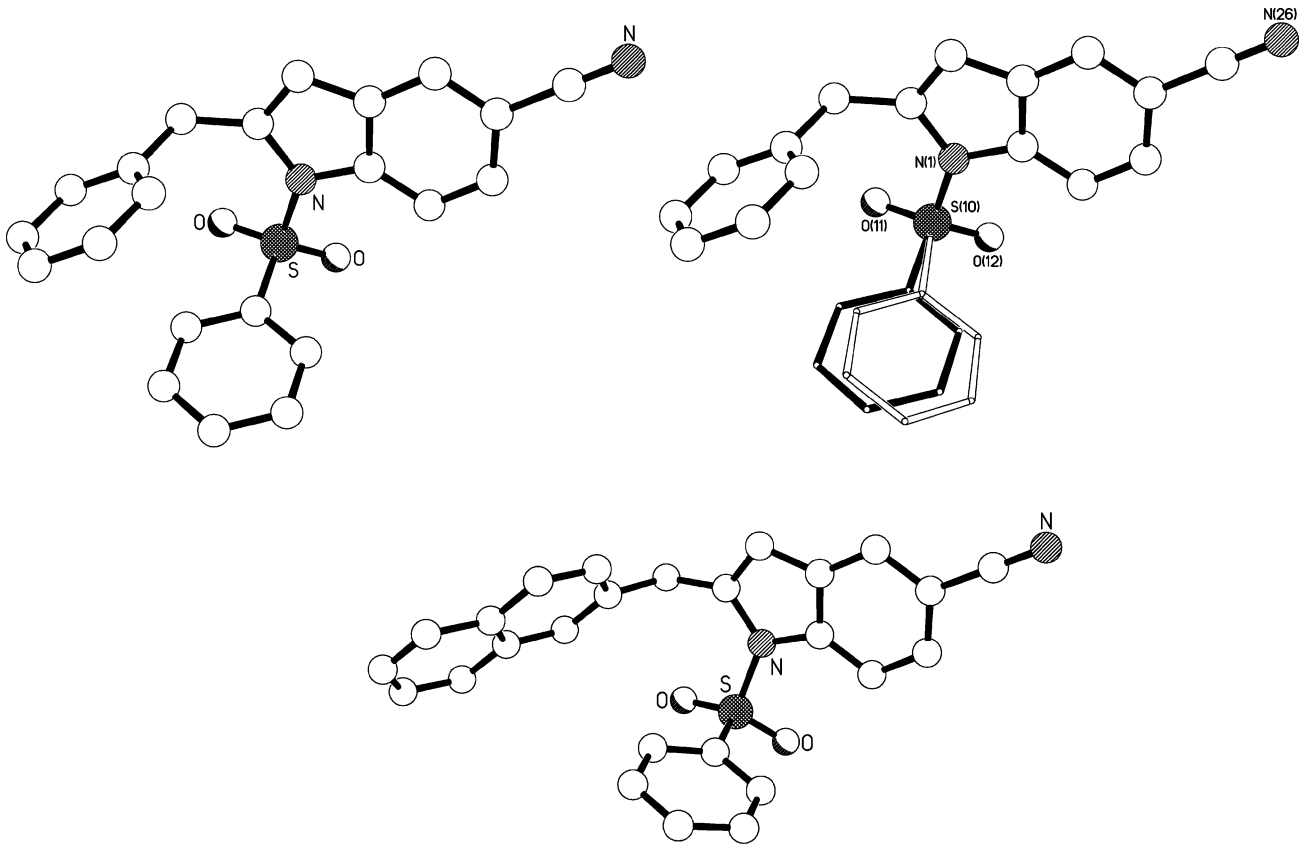


Fig. 5 The molecular structures of the 2-benzyl indole intermediate **22** (both top structures) and the molecular structure of the 2-naphthylmethyl intermediate **23** (bottom structure).

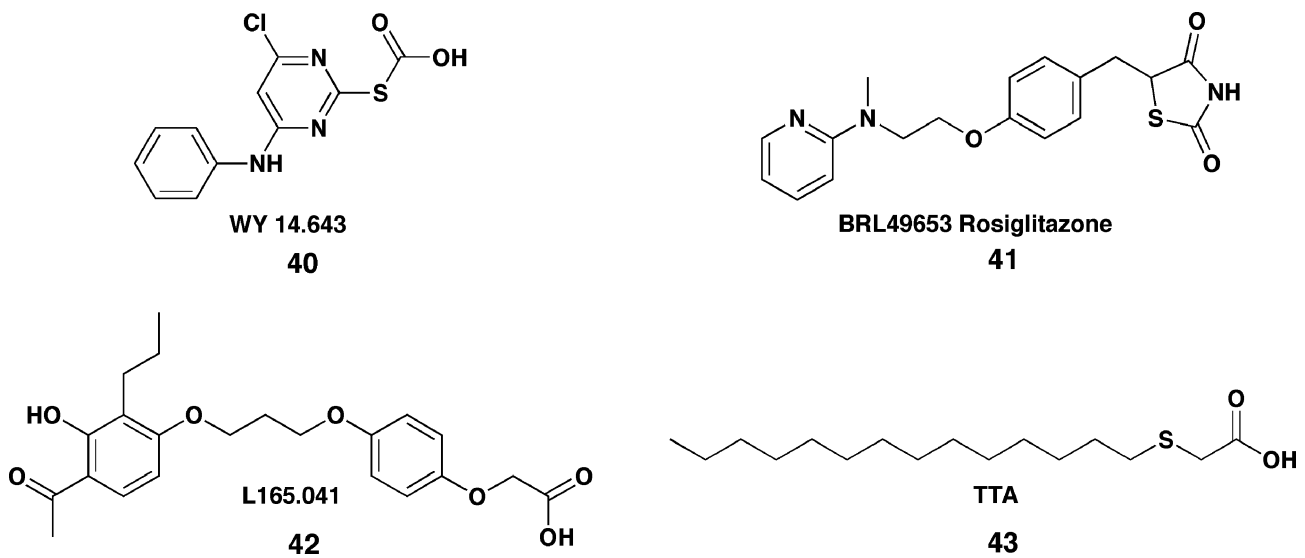


Fig. 6 Selective PPAR agonists used as positive controls and TTA **43**.

the best drug-like properties. Once the best candidate is chosen, the enantioselective synthesis can be carried out to give the desired enantiomer and to proceed to further testing of the compound as a potential treatment of metabolic diseases as well as inflammatory diseases. Thereafter we shall turn our attention to other novel chemotypes suggested by the software program PROTOBUILD.

Experimental

General chemistry

Unless otherwise stated all reactions were carried out under an atmosphere of nitrogen or argon, in oven-dried glassware. DCM was distilled over P_2O_5, and other solvents were bought and pre-dried

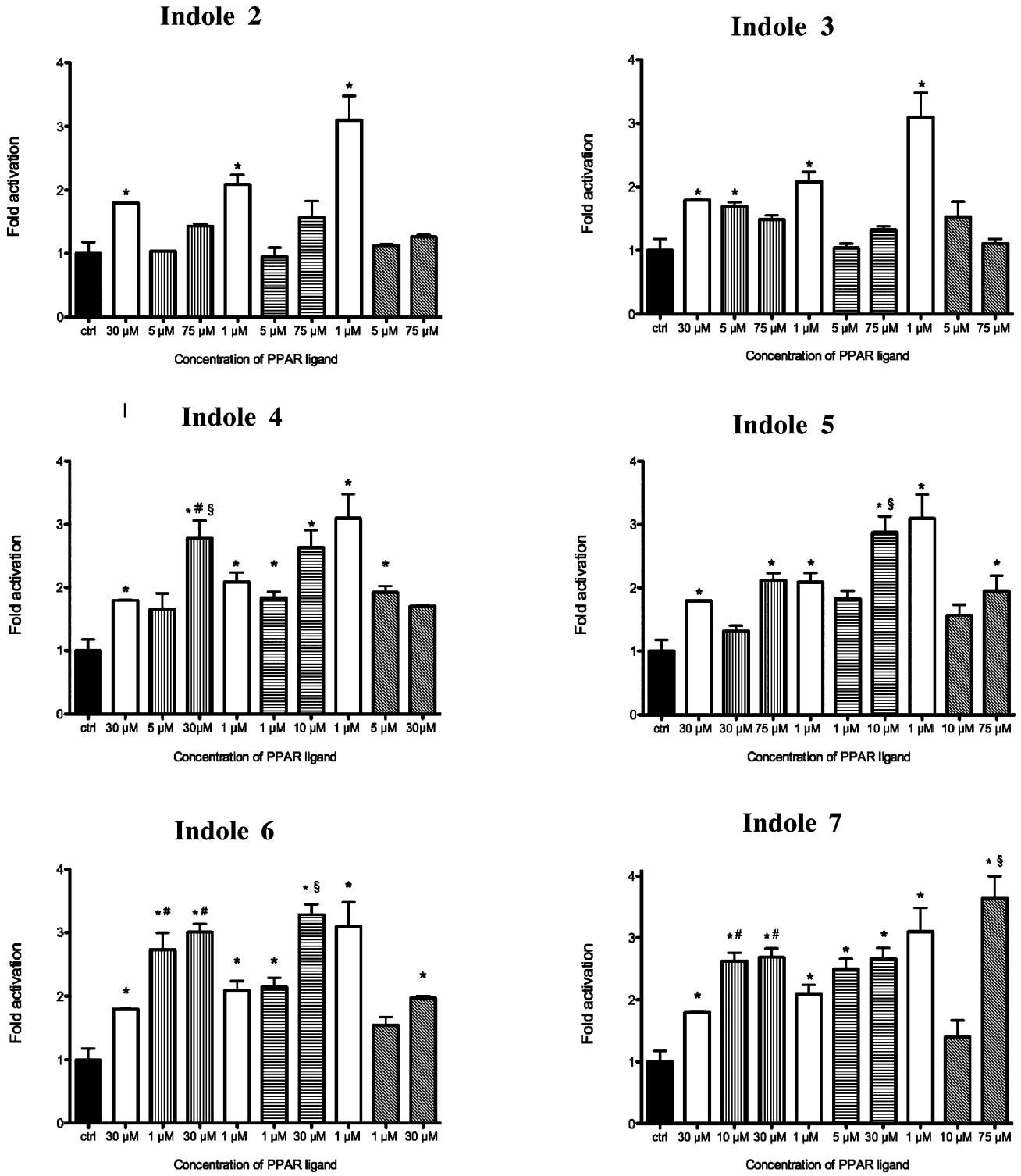


Fig. 7 PPAR activation with the indoles **2** to **7** in transfected human MCF-7 cells. For each indole, relative activation is given in the following order from left to right: control cells (black column), cells incubated with a PPARα agonist **40** (WY14.643), PPARγ agonist **41** (BRL49653) or a PPARδ agonist **42** (L165.041) (blank columns). In between is shown the activation of PPARα, PPARγ and PPARδ respectively with two different concentrations of the specific indole compound (grey-lined columns). Data are presented as mean ± SD ($n = 3$). *Significantly higher activation than in control cells ($P < 0.05$), #significantly higher activation than with the corresponding specific PPAR agonist ($P < 0.05$), §significantly higher activation than with lowest concentration of the same indole and same PPAR isoform ($P < 0.05$).

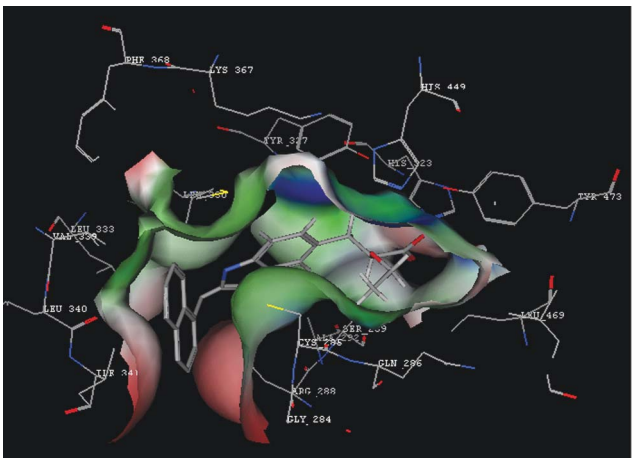


Fig. 8 Proposed interaction of indole **6** with the active site of PPARγ.

as required. All chemicals were purchased from Sigma-Aldrich, Lancaster or Merck Biosciences. *Flash column chromatography* (Merck Kieselgel 60 F_{254} 230-240 mesh) was performed according to the method of Still *et al.*[65] Thin layer chromatography (TLC) was performed on pre-coated Merck silica gel (0.2 mm, 60 F_{254}) aluminium-backed plates, and visualised with a UV lamp (254 nm) and/or stained with acidic ammonium molybdate (IV), basic potassium manganate (VII; $KMnO_4$), iodine or phosphomolybdic acid. Chromatography solvent mixture A is as follows: CH_2Cl_2–MeOH–H_2O 65 : 25 : 4. *Melting points* were measured on a Stuart Scientific SMP3 apparatus and are reported without correction. *Infrared spectra* (IR) were measured on a JASCO FT/IR-620 spectrometer. *^1H NMR* and *^{13}C NMR spectra* were recorded on Bruker Avance 400. ^{1}H NMR was recorded at 400 MHz using residual isotopic solvent as internal reference ($CDCl_3$, δ_H = 7.27 ppm; CD_3OD, δ_H = 3.30 ppm). ^{13}C NMR spectra were recorded at 100 MHz using residual isotopic solvent as internal reference ($CDCl_3$, δ_C = 77.00 ppm; CD_3OD, δ_C = 49.05 ppm). Peaks split by the presence of a phosphorus atom are indicated with a superscript p. *Mass spectra* were recorded using VG Platform II, VG-070B, Joel SX-102 or Bruker Esquire 3000 ESI instruments. Mass accuracy is indicated to the nearest 0.1 ppm. *Elemental analysis* was carried out using a Perkin Elmer 2400 CHN elemental analyser at the Science Technical Support unit, London Metropolitan University.

Ethyl 2-(diethoxyphosphoryl)-2-ethoxyacetate 9. This synthesis was performed in accordance to Grell and Machleidt.[57] In brief, triethylphosphite (1.46 mL, 8.49 mmol) was added to a colourless solution of 2-chloro-2-ethoxyacetic acid ethyl ester **17** (1.41 g, 8.49 mmol) in anhydrous DMF (10 mL) and the reaction mixture refluxed for 15 h at 145 °C. The resulting yellow solution was cooled to ambient temperature and water (40 mL) added. The aqueous phase was extracted with DCM, dried over $MgSO_4$ and concentrated *in vacuo* yielding 2.11 g of **9** as a pale yellow oil (92%): R_f 0.25 [hexane–EtOAc 1 : 1]; ν_{max}(film)/cm^{-1} 1746, 1265, 1145, 1039; ^{13}C NMR ($CDCl_3$): δ 167.33^p (*C*, *J* 2.0 Hz), 76.55^p (*C*H, *J* 157.0 Hz), 68.29^p (*C*H$_2$, *J* 12.0 Hz), 63.43^p (*C*H$_2$, *J* 7.0 Hz), 63.34 (*C*H$_2$, *J* 7.0 Hz), 61.45 (*C*H$_2$), 16.12^p (*C*H$_3$, *J* 6.0 Hz), 14.70 (*C*H$_3$), 13.88 (*C*H$_3$); ^{1}H NMR ($CDCl_3$): δ 4.24–4.09 (5H, m, H7, H5), 3.68–3.61 (1H, m, H3), 3.54–3.47 (1H, m, H3), 1.29–1.21 (9H, m, H8, H4 or H6), 1.19–1.16 (3H, m, H4 or H6); *m/z* (EI$^+$) 269 ([M+H]$^+$, 10), 224 ([M–ethyl ether]$^+$, 81), 197 ([M–(ether+ester)]$^+$, 90), 167 (85), 152 (70), 139 (75), 111 (100), 103 (55), 75 (62), 65 (71); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 268.1076 for $C_{10}H_{21}O_6P$, found 268.1302; Elemental calcd C, 44.8; H, 7.9; found 44.6; H, 7.9.

(*Z*)-Ethyl 2-ethoxy-3-(1*H*-indol-5-yl)acrylate 11 and (*E*)-ethyl 2-ethoxy-3-(1*H*-indol-5-yl)acrylate 12. Procedure **A**: This synthesis was performed in accordance to Kuhn *et al.*[51] In brief, phosphonoacetate **9** (1.87 g, 6.97 mmol) in THF (2.0 mL) was slowly added to an ice cold suspension of sodium hydride (60% in oil; 310 mg, 7.75 mmol) with molecular sieves in THF (40 mL), then stirred at 0 °C for 1 h. 5-indole-carbaldehyde **8** (506 mg, 3.48 mmol) in THF (4.0 mL) was added dropwise and the resulting orange solution stirred at 0 °C for 1 h and at ambient temperature for 44 h. After concentration *in vacuo*, the material was re-dissolved in EtOAc (40 mL) and washed twice with water, with the aqueous phase being extracted with EtOAc After drying over $MgSO_4$, concentration *in vacuo* gave a brown liquid which was purified by flash column chromatography (hexane–EtOAc 20 : 1, 15 : 1, 10 : 1, 9 : 1, 8 : 2) to yield 844 mg (93%) of two isomers. Ratio *Z* : *E* = 63 : 37. *Z* isomer **11**: amorphous white powder; mp 77.5–78.5 °C (hexane–DCM 20 : 1); R_f 0.35 [hexane–EtOAc 6 : 4]; ν_{max} (nujol)/cm^{-1} 3267, 2923–2850, 1731, 1687, 1550, 1378; ^{13}C NMR ($CDCl_3$): δ 165.30 (*C*), 142.56, 136.03, 127.98 (*C*), 126.19 (*C*H), 125.60 (*C*), 124.90 (*C*H), 124.56 (*C*H), 123.46 (*C*H), 111.00 (*C*H), 103.36 (*C*H), 67.46 (*C*H$_2$), 60.92 (*C*H$_2$), 15.55, 14.35 (*C*H$_3$); ^{1}H NMR ($CDCl_3$): δ 8.32 (1H, br, H1), 8.13 (1H, s, H4), 7.73 (1H, dd, *J* 8.4, 1.6 Hz, H6), 7.39 (1H, d, *J* 8.8 Hz, H7), 7.23 (1H, t, *J* 2.8 Hz, H2), 7.20 (1H, s, H1′), 6.60–6.58 (1H, m, H3), 4.32 (2H, q, *J* 7.2 Hz, H3′), 4.03 (2H, q, *J* 7.0 Hz, H6′), 1.44–1.38 (6H, m, H4′, H7′); *m/z* (EI$^+$) 259 ([M]$^+$, 100), 202 ([M–CH_2CH_3-CH_2CH_3]$^+$, 39), 157 (75), 129 (51), 118 ([Indole]$^+$, 13); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 259.1208 for $C_{15}H_{17}NO_3$, found 259.1199; Elemental calcd C, 69.5; H, 6.6; N, 5.4; found C, 69.6; H, 6.7; N, 5.3. *E* isomer **12**: amorphous white powder; mp 98.5–100.5 °C (hexane–DCM 20 : 1); R_f 0.28 [hexane–EtOAc 6 : 4]; ν_{max} (nujol)/cm^{-1} 3330, 2923–2850, 1702, 1621, 1379, 1227, 1176; ^{13}C NMR ($CDCl_3$): δ 165.24 (*C*), 146.23, 134.93, 127.82, 126.17 (*C*), 124.54 (*C*H), 122.99 (*C*H), 120.54 (*C*H), 111.47 (*C*H), 110.58 (*C*H), 102.57 (*C*H), 64.62 (*C*H$_2$), 61.07 (*C*H$_2$), 14.53 (*C*H$_3$), 13.68 (*C*H$_3$); ^{1}H NMR ($CDCl_3$): δ 8.23 (1H, br, H1), 7.51 (1H, d, *J* 0.4 Hz, H4), 7.30 (1H, d, *J* 8.4 Hz, H7), 7.17 (1H, t, *J* 2.8 Hz, H2), 7.07 (1H, dd, *J* 8.4, 1.6 Hz, H6), 6.50–6.49 (1H, m, H3), 6.32 (1H, s, H1′), 4.16 (2H, q, *J* 7.0 Hz, H3′), 3.96 (2H, q, *J* 6.8 Hz, H6′), 1.43 (3H, t, *J* 6.8 Hz, H7′), 1.08 (3H, t, *J* 7.0 Hz, H4′); *m/z* (EI$^+$) 259 ([M]$^+$, 100), 202 ([M–CH_2CH_3-CH_2CH_3]$^+$, 37), 157 (68), 129 (49); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 259.1208 for $C_{15}H_{17}NO_3$, found 259.1199; Elemental calcd C, 69.5; H, 6.6; N, 5.4; found C, 69.6; H, 6.5; N, 5.5.

Methyl 2-ethoxy-3-(1*H*-indol-5-yl)propanoate 15. Procedure **B**: Magnesium turnings (648 mg, 26.7 mmol, 18.2 eq) were added to a solution of **11** (380 mg, 1.47 mmol, 1.00 eq) in anhydrous methanol (13.0 ml), under a constant flow of N_2 at ambient temperature. The reaction mixture was stirred for 3 h then poured onto water and the aqueous phase extracted with DCM. DCM extracts were dried over $MgSO_4$, then concentrated *in vacuo* to give a residue that was purified by flash column chromatography (hexane–EtOAc 9 : 1, 8 : 2, EtOAc) to yield 300 mg (83%) of **15** as a yellow oil: R_f 0.20 [hexane–EtOAc 7.5 : 2.5]; ν_{max}(film)/cm^{-1} 3408, 2977–2850, 1741, 1444, 1209, 1110; ^{13}C NMR ($CDCl_3$): δ 173.30

(*C*), 134.80 (*C*), 128.19 (*C*), 127.98 (*C*), 124.38 (*C*H), 123.54 (*C*H), 121.04 (*C*H), 110.74 (*C*H), 102.23 (*C*H), 80.98 (*C*H), 66.21 (CH_2), 51.73 (CH_3), 39.51 (CH_2), 15.02 (CH_3); ^{1}H NMR ($CDCl_3$): δ 8.25 (1H, br, H1), 7.51 (1H, s, H4), 7.31 (1H, d, *J* 8.0 Hz, H7), 7.18 (1H, t, *J* 2.8 Hz, H2), 7.10 (1H, dd, *J* 8.4, 1.6 Hz, H6), 6.51–6.50 (1H, m, H3), 4.13 (1H, dd, *J* 7.6, 1.6 Hz, H2′), 3.71 (3H, s, H4″), 3.65–3.58 (1H, m, H3′), 3.42–3.35 (1H, m, H3′), 3.15 (1H, s, H1′), 3.13 (1H, d, *J* 2.8 Hz, H1′), 1.18 (3H, t, *J* 7.0 Hz, H4′); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 247.1208 for $C_{14}H_{17}NO_3$, found 247.1205; Elemental calcd C, 68.0; H, 6.9; N, 5.7; found C, 68.2; H, 6.9; N, 5.8.

2-Ethoxy-3-(1*H*-indol-5-yl)propanoic acid 2. Procedure **C**: Potassium hydroxide (15.1 mg, 0.269 mmol, 1.25 eq) was added to a solution of **15** (53.4 mg, 0.216 mmol, 1.00 eq) dissolved in a 1 : 1 mixture of ethanol and water, and the mixture was then stirred at 78 °C for 15 h. Thereafter the whole was concentrated *in vacuo*, and the concentrate subjected to vigorous stirring at 0 °C as the solution pH was adjusted to 1 with additions of aliquots of 1 M HCl. The resulting precipitate was dissolved in DCM and this organic solution was washed with water. Subsequently, the combined aqueous washes were re-extracted with DCM, and the combined DCM extracts were dried over $MgSO_4$ then concentrated *in vacuo* to give 46.2 mg (92%) of **2** as a pale yellow oil: R_f 0.45 [Solvent A]; ν_{max} (DCM film)/cm^{-1} 3411, 2978–2929, 1724, 1265, 1105; ^{13}C NMR ($CDCl_3$): δ 177.33 (*C*), 134.76 (*C*), 127.84 (*C*), 127.49 (*C*), 124.63 (*C*H), 123.25 (*C*H), 120.95 (*C*H), 110.93 (*C*H), 101.80 (*C*H), 80.27 (*C*H), 66.52 (CH_2), 38.91 (CH_2), 14.82 (CH_3); ^{1}H NMR ($CDCl_3$): δ 10.8 (1H, br, O*H*), 8.47 (1H, br, H1), 7.57 (1H, s, H4), 7.29 (1H, d, *J* 8.4 Hz, H7), 7.15–7.13 (2H, m, H2, H6), 6.52 (1H, br, H3), 4.22 (1H, dd, *J* 7.8, 4.6 Hz, H2′), 3.70–3.62 (1H, m, H3′), 3.49–3.41 (1H, m, H3′), 3.29 (1H, dd, *J* 14.2, 4.2 Hz, H1′), 3.19 (1H, dd, *J* 14.0, 7.6 Hz, H1′), 1.19 (3H, t, *J* 7.0 Hz, H4′); *m/z* (EI$^+$) 233 ([M]$^+$, 22), 130 (M–(ethoxypropanoic acid side chain)]$^+$, 100); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 233.1052 for $C_{13}H_{15}NO_3$, found 233.1047; Elemental calcd C, 66.9; H, 6.5; N, 6.0; found C, 67.0; H, 6.4; N, 5.9.

Ethyl 2-diazo-2-(diethoxyphosphoryl)acetate 18. The formation of tosyl azide was accomplished according to the published procedure procedure.[58] In brief, *p*-toluene sulfonyl chloride (14.9 g, 78.2 mmol, 1.00 eq) in acetone (70 ml) and sodium azide (5.10 g, 78.4 mmol, 1.00 eq) in ethanol (25 ml) were combined yielding 15.1 g of the desired reagent as a colourless liquid (98%) R_f 0.40 [hexane–EtOAc 8 : 2]; ^{1}H NMR ($CDCl_3$): δ 7.85–7.83 (2H, m, H2), 7.42–7.40 (2H, m, H3), 2.48 (3H, br, H1′); *m/z* (EI$^+$) 197 ([M]$^+$, 8), 155 ([M–azide]$^+$, 70), 91 ([methylbenzene]$^+$, 100); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 197.0259 for $C_7H_7N_3O_2S$, found 197.0256; Elemental calcd C, 42.6; H, 3.6; N, 21.3; found C, 42.7; H, 3.6; N, 21.3. This next stage was accomplished according to the procedure of Regitz *et al.*[60] In brief triethylphosphonoacetate **19** (0.63 ml, 3.20 mmol, 1.00 eq) was added dropwise to an ice-cold suspension of sodium hydride (157 mg, 3.92 mmol, 1.23 eq) in dry THF (8.00 ml). The reaction mixture was stirred at 0 °C for 45 min after which time tosyl azide (630 mg, 3.20 mmol, 1.00 eq) in dry THF (2 × 1.00 ml) was added, under a constant flow of N_2. The reaction mixture was further stirred for 16 h at 0 °C then quenched by the addition of ether and water. Thereafter, the aqueous phase was extracted with ether and the combined ethereal extracts were then washed with aq NaOH (0.5 M), water and brine. Finally the aqueous phase was extracted with ether and the combined organic extracts were then dried over $MgSO_4$ and concentrated *in vacuo*. The residue was purified *via* flash column chromatography (hexane, hexane–EtOAc 8 : 2, 6 : 4), yielding 586 mg (73%) of **18** as a pale yellow liquid: R_f 0.20 [hexane–EtOAc 1 : 1]; ν_{max}(film)/cm^{-1} 2987, 2131, 1707, 1284, 1026; ^{13}C NMR ($CDCl_3$): δ 163.34^p (*C*, d, *J* 12.0 Hz), 127.90^p (*C*, d, *J* 316.0 Hz), 63.56^p (CH_2, d, *J* 6.0 Hz), 61.61 (CH_2), 16.06 (CH_3, d, *J* 7.0 Hz), 14.25 (CH_3); ^{1}H NMR ($CDCl_3$): δ 4.25 (2H, q, *J* 7.2 Hz, H5), 4.23–4.10 (4H, m, H3), 1.34 (6H, dt, *J* 7.1, 0.7 Hz, H4), 1.29 (3H, t, *J* 7.2 Hz, H6); *m/z* (EI$^+$) 250 ([M]$^+$, 16), 121 (100), 109 (98), 93 (73), 81 (68), 65 (87); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 250.0719 for $C_8H_{15}N_2O_5P$ found 250.0711; Elemental calcd C, 38.4; H, 6.0; N, 11.2; found C, 38.5; H, 6.0; N, 11.1.

Ethyl 2-(diethoxyphosphoryl)-2-(2,2,2-trifluoroethoxy)acetate 10. This conversion was performed according to the method of Haigh *et al.*[66] In brief, a mixture of diazo ester **18** (586 mg, 2.34 mmol, 1.00 eq), trifluoroethanol (0.340 ml, 4.68 mmol, 2.00 eq) and Rh(II) acetate dimer (10.5 mg, 0.024 mmol, 0.01 eq) in benzene (7.00 ml) was heated with stirring at 85 °C for 22 h. Thereafter, the solution was concentrated *in vacuo* to a residue that was purified by flash column chromatography (hexane–EtOAc 8 : 2, 7 : 3, 6 : 4, 1 : 1) to yield 474 mg (63%) of **10** as a colourless liquid: R_f 0.20 [hexane–EtOAc 4 : 6]; ν_{max}(film)/cm^{-1} 1743, 1265, 1041; ^{13}C NMR ($CDCl_3$): δ 166.09 (*C*), 123.35 (*C*, q, *J* 277.5 Hz), 77.11^p (*C*H, d, *J* 156.0 Hz), 68.52^p (CH_2, dq, *J* 35.0, 11.5 Hz), 63.99^p (CH_2, t, *J* 6.0 Hz), 62.21 (CH_2), 16.23^p (CH_3, d, *J* 6.0 Hz), 13.99 (CH_3); ^{1}H NMR ($CDCl_3$): δ 4.50 (1H, d, *J* 14.4, H2), 4.37–4.08 (7H, m, H5, H7, H3), 3.98–3.89 (1H, m, H3), 1.36–1.30 (9H, m, H6, H8); *m/z* (EI$^+$) 322 ([M]$^+$, 1), 249 ([M–ethyl]$^+$, 20), 183 ([M–(trifluoroether + 2 ethyl groups)]$^+$, 61), 155 ([M–(trifluoroether + 3 ethyl groups)]$^+$, 100), 65 (64); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 322.0793 for $C_{10}H_{18}F_3O_6P$ found 322.0787; Elemental calcd C, 37.3; H, 5.6; found C, 37.3; H, 5.6.

(*Z*)-Ethyl 3-(1*H*-indol-5-yl)-2-(2,2,2-trifluoroethoxy)acrylate 13 and (*E*)-ethyl 3-(1*H*-indol-5-yl)-2-(2,2,2-trifluoroethoxy)acrylate 14. Procedure **D**: phosphonoacetate **10** (60.0 mg, 0.186 mmol, 1.06 eq) in THF (0.600 ml) was added dropwise to a suspension of sodium hydride (9.80 mg, 0.245 mmol, 1.40 eq) in anhydrous THF (0.800 ml) at 0 °C under a constant flow of N_2. Thereafter the reaction mixture was stirred at 0 °C for 25 min, then 5-indolecarbaldehyde **8** (25.4 mg, 0.175 mmol, 1.00 eq) was added in dry THF (0.900 ml). Post stirring for a further 25.5 h at 0 °C, the solution was concentrated *in vacuo* and the residue dissolved in EtOAc for washing with water. Combined aqueous washes were extracted with EtOAc and the combined organic extracts were then washed with brine, dried over $MgSO_4$, concentrated *in vacuo* leaving a residue that was purified by flash column chromatography (hexane–EtOAc 20 : 1, 15 : 1, 10 : 1, 8 : 2) to yield 54.3 mg (99%) of two isomers. Ratio *Z* : *E* = 59 : 41. *Z* isomer **13**: amorphous white powder; mp 98.0–99.0 °C (hexane–DCM); R_f 0.20 [hexane–EtOAc 7.5 : 2.5]; ν_{max} (nujol)/cm^{-1} 3343, 2977–2850, 1691, 1459, 1371, 1261, 1056; ^{13}C NMR ($CDCl_3$): δ 164.05 (*C*), 140.54 (*C*), 136.37 (*C*), 128.08 (*C*), 123.43 (*C*, q, *J* 277.0 Hz), 127.34 (*C*H), 125.06 (*C*H), 124.84 (*C*H), 124.84 (C), 124.19 (*C*H), 111.16 (*C*H), 103.62 (*C*H), 67.86 (CH_2, q, *J* 34.5 Hz), 61.34 (CH_2), 14.29 (CH_3); ^{1}H NMR ($CDCl_3$): δ 8.26 (1H, br, H1), 8.11 (1H, s, H4), 7.68 (1H, dd, *J* 8.6, 1.4 Hz, H6), 7.41 (1H, d, *J* 8.8 Hz, H7), 7.27 (1H, br, H1′), 7.24 (1H, t, *J* 2.8, H2), 6.61–6.60 (1H,

m, H3), 4.41–4.31 (4H, m, H3′, H6′), 1.40 (3H, t, *J* 7.2 Hz, H7′); *m/z* (EI$^+$) 313 ([M]$^+$, 100), 230 ([M–CH_2CF_3]$^+$, 6), 202 ([M–CH_2CH_3-CH_2CF_3]$^+$, 39), 157 (33), 129 (44), 118 ([Indole]$^+$, 13); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 313.0926 for $C_{15}H_{14}F_3NO_3$ found 313.0925; Elemental calcd C, 57.7; H, 4.5; N, 4.5; found C, 57.6; H, 4.6; N, 4.5. *E* isomer **14**: amorphous pale yellow powder; mp 88.0–89.0 °C (hexane–DCM); R_f 0.15 [hexane–EtOAc 7.5 : 2.5]; ν_{max} (nujol)/cm^{-1} 3372, 2950–2850, 1711, 1461, 1378, 1269, 1157; ^{13}C NMR ($CDCl_3$): δ 163.51(*C*), 143.55 (*C*), 135.52 (*C*), 127.76 (*C*), 124.77 (*C*H), 124.54 (*C*), 123.42 (*C*H), 123.24 (*C*, q, *J* 277.0 Hz), 121.67 (*C*H), 121.56 (*C*H), 110.62 (*C*H), 103.00 (*C*H), 68.16 (*C*H_2, q, *J* 35.0 Hz), 61.29 (*C*H_2), 13.75 (*C*H_3); ^{1}H NMR ($CDCl_3$): δ 8.19 (1H, br, H1), 7.62 (1H, d, *J* 0.8 Hz, H4), 7.34 (1H, d, *J* 8.4 Hz, H7), 7.21 (1H, t, *J* 2.8 Hz, H2), 7.15 (1H, dd, *J* 8.4, 1.6 Hz, H6), 6.74 (1H, s, H1′), 6.55–6.53 (1H, m, H3), 4.27 (2H, q, *J* 8.4 Hz, H3′), 4.20 (2H, q, *J* 7.2 Hz, H6′), 1.14 (3H, t, *J* 7.2 Hz, H7′); *m/z* (EI$^+$) 313 ([M]$^+$, 100), 230 ([M–CH_2CF_3]$^+$,6), 202 ([M–CH_2CH_3-CH_2CF_3]$^+$, 38), 157 (34), 129 (48), 118 ([Indole]$^+$, 15); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 313.0926 for $C_{15}H_{14}F_3NO_3$ found 313.0926; Elemental calcd C, 57.7; H, 4.5; N, 4.5; found C, 57.6; H, 4.6; N, 4.4.

Methyl 3-(1*H*-indol-5-yl)-2-(2,2,2-trifluoroethoxy)propanoate 16. This transformation was carried out as described in procedure **B** using starting material **13**. After purification *via* flash column chromatography (hexane–EtOAc 20 : 1, 10 : 1, 9 : 1, 8 : 2) **16** was obtained as a pale yellow oil 653 mg (60%): R_f 0.20 [hexane–EtOAc 7.5 : 2.5]; ν_{max} (DCM film)/cm^{-1} 3377, 2933, 1741, 1513, 1278, 1164; ^{13}C NMR ($CDCl_3$): δ 171.50 (*C*), 134.91 (*C*), 128.05 (*C*), 123.56 (C, q, *J* 277.0 Hz), 127.42 (*C*), 124.43 (*C*H), 123.62 (*C*H), 121.25 (*C*H), 110.84 (*C*H), 102.47 (*C*H), 82.02 (*C*H), 67.89 (*C*H_2, q, *J* 34.0 Hz), 52.11 (*C*H_2), 39.17 (*C*H_2); ^{1}H NMR ($CDCl_3$): δ 8.13 (1H, br, H1), 7.51 (1H, s, H4), 7.33 (1H, d, *J* 8.4 Hz, H7), 7.20 (1H, t, *J* 2.8 Hz, H2), 7.09 (1H, dd, *J* 8.4, 1.6 Hz, H6), 6.52–6.51 (1H, m, H3), 4.26 (1H, dd, *J* 7.6, *J* 4.8 Hz, H2′), 4.02–3.93 (1H, m, H3′), 3.75 (3H, s, H6′), 3.73–3.66 (1H, m, H3′), 3.24 (1H, dd, *J* 14.0 4.8 Hz, H1′), 3.17 (1H, dd, *J* 14.2, 7.8 Hz, H1′); *m/z* (EI$^+$) 301 ([M]$^+$, 100), 242 (65), 201 ([M–CH_2CH_3-CH_2CF_3]$^+$, 39), 131 (80); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 301.0926 for $C_{14}H_{14}F_3NO_3$ found 301.0923; Elemental calcd C, 55.8; H, 4.7; N, 4.7; found C, 55.9; H, 4.7; N, 4.6.

3-(1*H*-indol-5-yl)-2-(2,2,2-trifluoroethoxy)propanoic acid 3. The reaction was performed according to procedure **C**, using starting material **16**. This process afforded 38.1 mg (93%) of **3** as a pale yellow oil: R_f 0.25 [Solvent A]; ν_{max} (DCM film)/cm^{-1} 3415, 2933, 1726, 1279, 1167; ^{13}C NMR ($CDCl_3$/MeOD 2 : 1): δ 173.97 (*C*), 135.68 (*C*), 128.53 (*C*), 124.18 (*C*, q, *J* 277.0 Hz), 127.67 (*C*H), 125.20 (*C*H), 123.52 (*C*H), 121.31 (*C*H), 111.40 (*C*H), 101.73 (*C*H), 82.44 (*C*H), 68.18 (*C*H_2, q, *J* 34.5 Hz), 39.63 (*C*H_2); ^{1}H NMR ($CDCl_3$/MeOD 2 : 1): δ 9.85 (1H, br, OH), 7.45 (1H, s, H4), 7.28 (1H, d, *J* 8.0 Hz, H7), 7.14–7.13 (1H, m, H2), 7.03 (1H, dd, *J* 8.4, 1.6 Hz, H6), 6.38–6.37 (1H, m, H3), 4.19 (1H, dd, *J* 8.4, 4.0 Hz, H2′), 3.97–3.87 (1H, m, H3′), 3.70–3.60 (1H, m, H3′), 3.18 (1H, dd, *J* 14.2, 4.2 Hz, H1′), 3.06 (1H, dd, *J* 14.2, 8.2 Hz, H1′); *m/z* (EI$^+$) 287 ([M]$^+$, 18), 130 ([M–(ethoxypropanoic acid side chain)]$^+$, 100); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 287.0769 for $C_{13}H_{12}NO_3F_3$ found 287.0763; Elemental calcd C, 54.4; H, 4.2; N, 4.9; found C, 54.7; H, 4.2; N, 4.8.

1-Benzenesulfonyl-1*H*-indole-5-carbonitrile 21. This reaction was performed according to a procedure of Iwanowicz *et al.*[67] In brief, a solution of benzenesulfonyl chloride (0.850 ml, 0.688 mmol, 1.10 eq) in toluene (4.00 ml) was added dropwise to a vigorously stirred biphasic system of indole-5-carbonitrile **20** (864 mg, 6.08 mmol, 1.00 eq) and *n*-tetrabutylammonium bromide (197 mg, 0.609 mmol, 0.10 eq) in 50% aqueous NaOH (5.90 ml), toluene (5.00 ml) and water (9.00 ml). Post 25 h stirring at ambient temperature, the aqueous phase was removed and the organic phase washed with 0.1 M $NaHCO_3$, water and saturated brine. All aqueous extracts were extracted with EtOAc then combined organic extracts were dried over $MgSO_4$ and concentrated *in vacuo* to give an off-white powder that was purified *via* recrystallisation from hexane–DCM 20 : 1 to yield 1.65 g (96%) of **21** as white needles: mp 131.0–133.5 °C (hexane–DCM 20 : 1); R_f 0.35 [hexane–EtOAc 6 : 4]; ν_{max} (nujol)/cm^{-1} 2222, 1458, 1375, 1080, 721; ^{13}C NMR ($CDCl_3$): δ 137.73 (*C*), 136.43 (*C*), 134.42 (*C*H), 130.65 (*C*), 129.54 (*C*H), 128.37 (*C*H), 127.59 (*C*H), 126.77 (*C*H), 126.36 (*C*H), 119.18 (*C*), 114.27 (*C*H), 108.65 (*C*H), 107.03 (*C*); ^{1}H NMR ($CDCl_3$): δ 8.10–8.08 (1H, d, *J* 8.4 Hz, H2′), 7.90–7.89 (3H, m, H2′, H4, H7), 7.71 (1H, d, *J* 3.6 Hz, H2), 7.62–7.56 (2H, m, H3′), 7.49 (2H, t, *J* 7.8 Hz, H4′, H6), 6.74 (1H, d, *J* 3.6 Hz, H3); *m/z* (EI$^+$) 282 ([M]$^+$, 50), 141 ([M–BzSulfonyl group]$^+$, 66), 114 ([unsubstituted indole]$^+$, 12), 77 (100); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 282.0463 for $C_{15}H_{10}N_2O_2S$ found 282.0474; Elemental calcd C, 63.8; H, 3.6; N, 9.9; found C, 63.7; H, 3.5; N, 9.9.

2-Benzyl-1-(phenylsulfonyl)-1*H*-indole-5-carbonitrile 22. This reaction was performed by adaptation of a procedure of Iwanowicz *et al.*[67] In brief, a solution of 2 M LDA solution in pentane (0.24 ml, 0.470 mmol, 1.20 eq) was added dropwise to a solution of phenylsulfonyl indole **21** (109 mg, 0.39 mmol, 1.00 eq) in anhydrous THF (4 ml) at −78 °C, under a constant flow of N_2. The reaction mixture was stirred at −78 °C for 55 min, followed by the dropwise addition of a solution of benzyl bromide (70 μL, 0.585 mmol, 1.5 eq) in dry THF (2 ml). The reaction mixture was allowed to warm up to 0 °C with stirring over 50 min, then to ambient temperature over 15 h. Thereafter, the mixture was quenched with 5% citric acid and the organic phase washed with water, then aqueous extracts were extracted with ethylacetate. Combined organic extracts were dried over $MgSO_4$ and concentrated *in vacuo*, giving a residue that was purified *via* flash column chromatography (hexane–EtOAc 10 : 1) to yield 56.2 mg (66.5%) of the product as yellow oil. Crystallisation from hexane–DCM 20 : 1 yielded **22** as white needles: mp 130.5–131.5 °C (hexane–DCM 20 : 1); R_f 0.20 [hexane–EtOAc 8 : 2]; ν_{max} (nujol)/cm^{-1} 2220, 1460, 1371, 1053; ^{13}C NMR ($CDCl_3$): δ 143.63 (*C*), 138.98 (*C*), 138.59 (*C*), 136.99 (*C*), 134.19 (*C*H), 129.46 (*C*H), 129.38 (*C*), 129.33 (*C*H), 128.69 (*C*H), 127.18 (*C*H), 127.03 (*C*H), 126.39 (*C*H), 125.04 (*C*H), 119.32 (*C*), 115.35 (*C*H), 110.07 (*C*H), 107.10 (*C*), 35.12 (*C*H_2); ^{1}H NMR ($CDCl_3$): δ 8.28 (1H, d, *J* 8.8 Hz, H7), 7.71 (1H, s, H4), 7.66 (2H, d, *J* 7.6 Hz, H2′), 7.58 (1H, t, *J* 7.6 Hz, H4′), 7.54 (1H, dd, *J* 8.8, 1.2 Hz, H6), 7.42 (2H, t, *J* 8.8 Hz, H3′), 7.34–7.29 (3H, m, H8′, H9′), 7.20–7.18 (2H, m, H7′), 6.16 (1H, s, H3), 4.37 (2H, s, H5′); *m/z* (EI$^+$) 372 ([M]$^+$, 57), 284 ([M–Bn group]$^+$, 30), 230 ([M–Phsulfonyl group]$^+$, 100), 141 (11), 115 ([indole]$^+$, 5), 91 (12), 77 (47); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 372.0932 for $C_{22}H_{16}N_2O_2S$ found 372.0934; Elemental calcd C, 71.0; H, 4.3; N, 7.5; found C, 71.0; H, 4.2; N, 7.6. *Crystal data*

for **22**: $C_{22}H_{16}N_2O_2S$, M = 372.43, monoclinic, $P2_1/n$ (no. 14), a = 10.8591(3), b = 10.0425(3), c = 17.0542(4) Å, β = 100.888(3)°, V = 1826.32(9) Å^3, Z = 4, D_c = 1.354 g cm^{-3}, μ(Mo-Kα) = 0.197 mm^{-1}, T = 173 K, colourless blocks, Oxford Diffraction Xcalibur 3 diffractometer; 4231 independent measured reflections, F^2 refinement, R_1 = 0.044, wR_2 = 0.101, 2710 independent observed absorption-corrected reflections [$|F_o| > 4\sigma(|F_o|)$, $2\theta_{max}$ = 58°], 245 parameters. CCDC 667594.

2-Benzyl-1*H*-indole-5-carbonitrile 24. This transformation was performed according to procedure **E**, adapted from protocols of Iwanowicz *et al.*[67] In brief 2-benzyl indole **22** (141 mg, 0.379 mmol, 1.00 eq) was refluxed at 110 °C for 18 h in methanolic 2 M NaOH, then the reaction mixture was allowed to cool to ambient temperature before acidification to pH 2 using aliquots of 0.1 M HCl. The aqueous phase was extracted with DCM, then organic extracts were dried using $MgSO_4$, concentrated *in vacuo* leaving a residue that was purified *via* flash column chromatography (hexane, hexane–EtOAc 10 : 1, 9 : 1, 8 : 2) giving 87.0 mg (99%) of **24** as an off-white solid: mp 119.5–120.5 °C (hexane–DCM 20 : 1); R_f 0.25 [hexane–EtOAc 7 : 3]; ν_{max} (nujol)/cm^{-1} 3300, 3056–2850, 2229, 1373, 1056, 740; ^{13}C NMR ($CDCl_3$): δ 140.38 (*C*), 137.94 (*C*), 137.53 (*C*), 128.94 (*C*H), 128.82 (*C*H), 128.49 (*C*), 127.10 (*C*H), 125.40 (*C*H), 124.49 (*C*H), 120.82 (*C*), 111.26 (*C*H), 102.84 (*C*), 101.66 (*C*H), 34.59 (*C*H$_2$); ^{1}H NMR ($CDCl_3$): δ 8.12 (1H, br, H1), 7.88 (1H, s, H4), 7.38–7.25 (6H, m, H4′, H7, H5′, H6, H3′), 6.41 (1H, s, H3), 4.16 (2H, s, H5′); *m/z* (EI$^+$) 232 ([M]$^+$, 100), 231 (64), 155 (71), 91 (16), 84 (21), 49 (23), 43 (25); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 232.1000 for $C_{16}H_{12}N_2$ found 232.1002; Elemental calcd C, 82.7; H, 5.2; N, 12.1; found C, 82.9; H, 5.1; N, 11.9.

2-Benzyl-1*H*-indole-5-carbaldehyde 26. This reaction was carried out according to procedure **F**: in brief a DIBAL solution, 1.0 M in DCM (2.20 ml, 2.20 mmol, 1.24 eq) was added dropwise under a constant flow of N_2 to a solution of 2-benzyl-1*H*-indole-5-carbonitrile **24** (411 mg, 1.77 mmol, 1.00 eq) in anhydrous DCM (8.00 ml) at 0 °C. The reaction mixture was allowed to warm with stirring to ambient temperature over 16 h. Thereafter, the mixture was cooled to 0 °C and Rochelle solution (1.0 M, 17.0 ml) was added. Post 4.5 h of stirring at ambient temperature, the two phases were separated and the aqueous phase was extracted with DCM. Combined organic extracts were concentrated *in vacuo* and the resulting orange slurry was dispersed in further DCM for acidification with HCl (1 M, 2.90 ml). After stirring for 1.7 h at ambient temperature, the mixture was finally quenched using NaOH (2 M, 2.90 ml) and extracted with DCM. Combined organic extracts were washed with water and brine, then dried over $MgSO_4$ and concentrated *in vacuo* to give an orange powder that was purified by flash column chromatography (hexane–EtOAc 10 : 1, 9 : 1) to yield 332 mg (80%) of **26** as a pale orange amorphous powder: mp 126.5–127.5 °C (hexane–DCM 20 : 1); R_f 0.20 [hexane–EtOAc 7.5 : 2.5]; ν_{max} (nujol)/cm^{-1} 3324, 3056–2850, 1671, 1373, 1056, 739; ^{13}C NMR ($CDCl_3$): δ 192.47 (*C*H), 139.95 (*C*), 139.88 (*C*), 137.76 (*C*), 129.67 (*C*H), 128.87 (*C*H), 128.81 (*C*H), 128.54 (*C*), 127.00 (*C*H), 125.05 (*C*H), 122.10 (*C*H), 111.04 (*C*H), 102.58 (*C*H), 34.65 (*C*H$_2$); ^{1}H NMR ($CDCl_3$): δ 10.01 (1H, s, H1″), 8.08 (2H, br, H1, H4), 7.70 (1H, dd, *J* 8.6, 1.4 Hz, H6), 7.38–7.28 (6H, m, H4′, H7, H5′, H3′), 6.49 (1H, br, H3), 4.17 (2H, s, H5′); *m/z* (EI$^+$) 235 ([M]$^+$, 100), 206 ([M–aldehyde]$^+$, 27), 167 (30), 158 (42), 149 (30), 91 (14); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 235.0997 for $C_{16}H_{13}NO$ found 235.0997; Elemental calcd C, 81.7; H, 5.6; N, 6.0; found C, C, 81.6; H, 5.5; N, 5.9.

(*Z*)-Ethyl 3-(2-benzyl-1*H*-indol-5-yl)-2-ethoxyacrylate 28 and (*E*)-ethyl 3-(2-benzyl-1*H*-indol-5-yl)-2-ethoxyacrylate 29. This reaction was carried out according to procedure **A** using the starting materials **26** and **9**. After purification *via* flash column chromatography (hexane–EtOAc 15 : 1, 12 : 1) the product could be obtained as a colourless oil, yielding 66.0 mg (90%) of two isomers. Ratio $Z : E$ = 63 : 37. *Z* isomer **28**: colourless oil: R_f 0.20 [hexane–EtOAc 8 : 2]; ν_{max} (DCM film)/cm^{-1} 3441, 3055, 1743, 1646, 1051, 740; ^{13}C NMR ($CDCl_3$): δ 165.31 (*C*), 142.43 (*C*), 138.67 (*C*), 138.21 (*C*), 136.56 (*C*), 128.81 (*C*H), 128.77 (*C*H), 126.82 (*C*H), 126.30 (*C*H), 125.53 (*C*), 124.08 (*C*H), 122.71 (*C*H), 110.43 (*C*H), 101.73 (*C*H), 67.41 (*C*H$_2$), 60.87 (*C*H$_2$), 34.69 (*C*H$_2$), 15.56 (*C*H$_3$), 14.36 (*C*H$_3$); ^{1}H NMR ($CDCl_3$): δ 8.03 (1H, br, H4), 7.89 (1H, br, H1), 7.63 (1H, dd, *J* 8.4, 1.6 Hz, H6), 7.37–7.33 (2H, m, H4′), 7.30–7.22 (4H, m, H3′, H7, H5′), 7.17 (1H, s, H1″), 6.35 (1H, d, *J* 0.8 Hz, H3), 4.32 (2H, q, *J* 7.2 Hz, H6″), 4.14 (2H, s, H1′), 4.02 (2H, q, *J* 7.1 Hz, H3″), 1.42–1.37 (6H, m, H7″, H4″); *m/z* (EI$^+$) 349 ([M]$^+$, 100), 292 ([M–ethyl]$^+$, 17), 292 (32), 247 (36), 219 (26), 91 (50); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 349.1678 for $C_{22}H_{23}NO_3$ found 349.1672; Elemental calcd C, 75.6; H, 6.6; N, 4.0; found C, 75.6; H, 6.5; N, 3.9. *E* isomer **29**: pale yellow oil: R_f 0.15 [hexane–EtOAc 8 : 2]; ν_{max} (DCM film)/cm^{-1} 3441, 3055–2850, 1725, 1644, 1548, 1266, 1043, 738; ^{13}C NMR ($CDCl_3$): δ 165.20 (*C*), 146.16 (*C*), 138.41 (*C*), 138.16 (*C*), 135.44 (*C*), 128.81 (*C*H), 128.74 (*C*H), 126.76 (*C*H), 126.26 (*C*), 122.40 (*C*H), 119.96 (*C*H), 111.60 (*C*H), 110.00 (*C*H), 101.20 (*C*H), 64.63 (*C*H$_2$), 61.05 (*C*H$_2$), 34.73 (*C*H$_2$), 14.58 (*C*H$_3$), 13.77 (*C*H$_3$); ^{1}H NMR ($CDCl_3$): δ 7.75 (1H, br, H1), 7.40 (1H, s, H4), 7.36–7.32 (2H, m, H4′), 7.28–7.25 (3H, m, H3′, H5′), 7.15 (1H, d, *J* 8.4 Hz, H7), 6.98 (1H, dd, *J* 8.4, 1.2 Hz, H6), 6.29 (1H, s, H1″), 6.27 (1H, br, H3), 4.18–4.12 (4H, m, H6″, H1′), 3.96 (2H, q, *J* 6.9 Hz, H3″), 1.44–1.41 (3H, t, *J* 7.0 Hz, H4″), 1.09 (3H, t, *J* 7.0 Hz, H7″); *m/z* (EI$^+$) 349 ([M]$^+$, 16), 232 ([M–(benzyl+ethyl)]$^+$, 100), 155 (77), 84 (42), 49 (56); *m/z* HRMS (EI$^+$) [M+H]$^+$ calcd 349.1678 for $C_{22}H_{23}NO_3$ found 349.1671; Elemental calcd C, 75.6; H, 6.6; N, 4.0; found C, 75.7; H, 6.8; N, 3.9.

Methyl 3-(2-benzyl-1*H*-indol-5-yl)-2-ethoxypropanoate 36. The reaction was performed according to procedure **B** using **28**. After purification *via* flash column chromatography (hexane–EtOAc 15 : 1, 10 : 1) 40.6 mg (76%) of **36** was obtained as a pale yellow oil: R_f 0.25 [hexane–EtOAc 7.5 : 2.5]; ν_{max} (DCM film)/cm^{-1} 3392, 3027–2900, 1741, 1644, 1446, 1115, 705; ^{13}C NMR ($CDCl_3$): δ 173.28 (*C*), 138.51 (*C*), 138.03 (*C*), 135.28 (*C*), 128.83 (*C*), 128.80 (*C*H), 128.68 (*C*H), 128.24 (*C*), 126.69 (*C*H), 122.91 (*C*H), 120.44 (*C*H), 110.16 (*C*H), 100.86 (*C*H), 81.04 (*C*H), 66.19 (*C*H$_2$), 51.73 (*C*H$_2$), 39.55 (*C*H$_2$), 34.72 (*C*H$_2$), 15.06 (*C*H$_3$); ^{1}H NMR ($CDCl_3$): δ 7.77 (1H, br, H1), 7.41 (1H, br, H4), 7.36–7.32 (2H, m, H4′), 7.29–7.26 (3H, m, H3′, H5′), 7.16 (1H, d, *J* 8.4 Hz, H7), 7.01 (1H, dd, *J* 8.4, 1.8 Hz, H6), 6.28 (1H, s, H3), 4.12 (2H, s, H1′), 4.09 (1H, t, *J* 6.6 Hz, H2″), 3.71 (3H, s, H6″), 3.64–3.56 (1H, m, H3″), 3.41–3.34 (1H, m, H3″), 3.11–3.09 (2H, m, H1″), 1.18 (3H, t, *J* 6.8 Hz, H6″); *m/z* (EI$^+$) 337 ([M]$^+$, 37), 234 ([M–(benzyl + methyl groups)]$^+$, 81), 220 ([M–(benzyl+2 methyl groups)]$^+$,100), 91 (29); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 337.1668 for

$C_{21}H_{23}NO_3$ found 337.1674; Elemental calcd C, 74.8; H, 6.9; N, 4.2; found C, 79.9; H, 5.2.; N, 4.2.

3-(2-Benzyl-1*H*-indol-5-yl)-2-ethoxy-propanoic acid 4. The synthesis was performed according to procedure **C** using **36**. This resulted in 29.8 mg (93%) of **4** as a light yellow oil: R_f 0.20 [hexane–EtOAc 1 : 1]; ν_{max} (DCM film)/cm^{-1} 3620, 3398, 1710, 1646, 1461, 1108; ^{13}C NMR ($CDCl_3$): δ 175.40 (*C*), 138.45 (*C*), 138.15 (*C*), 135.35 (*C*), 128.86 (*C*), 128.81 and 128.70 (*C*H each), 127.69 (*C*), 126.71 (*C*H), 122.95 (*C*H), 120.63 (*C*H), 110.27 (*C*H), 100.89 (*C*H), 80.36 (*C*H), 66.82 (CH_2), 38.85 (CH_2), 34.71 (CH_2), 15.05 (CH_2); ^{1}H NMR ($CDCl_3$): δ 7.77 (1H, br, H1), 7.42 (1H, s, H4), 7.36–7.32 (2H, m, H4′), 7.29–7.25 (3H, m, H3′, H5′), 7.17 (1H, d, *J* 8.4 Hz, H7), 7.02 (1H, dd, *J* 8.2, 1.4 Hz, H6), 6.28 (1H, s, H3), 4.14–4.12 (3H, m, H2″, H1′), 3.63–3.56 (1H, m, H3″), 3.49–3.42 (1H, m, H3″), 3.22 (1H, dd, *J* 14.2, 4.2 Hz, H1″), 3.10 (1H, dd, *J* 14.2, 7.8 Hz, H1″), 1.18 (3H, t, *J* 7.0 Hz, H4″); *m*/*z* (EI$^+$) 323 ([M]$^+$, 16), 220 ([M–(benzyl+methyl group)]$^+$, 88), 199 (29), 171 (M–(benzyl + ether side chain + OH)]$^+$, 100), 127 (51), 57 (53); *m*/*z* HRMS (EI$^+$) [M]$^+$ calcd 323.1521 for $C_{20}H_{21}NO_3$ found 323.1520; Elemental calcd C, 74.3; H, 6.6; N, 4.3; found C, 74.2; H, 6.5; N, 4.4.

(*Z*)-Ethyl 3-(2-benzyl-1*H*-indol-5-yl)-2-(2,2,2-trifluoroethoxy)-acrylate 30 and (*E*)-ethyl 3-(2-benzyl-1*H*-indol-5-yl)-2-(2,2,2-trifluoroethoxy)acrylate 31. The synthesis was carried out according to general procedure **D** using the starting materials **26** and **10**. After purification *via* flash column chromatography (hexane–EtOAc 15 : 1, 10 : 1), 107 mg (85%) of two isomers was obtained. Ratio *Z* : *E* = 39 : 61. *Z* isomer **30**: yellow amorphous powder; mp 120.5–121.5 °C (hexane–DCM 20 : 1); R_f 0.30 [hexane–EtOAc 7.5 : 2.5]; ν_{max} (DCM film)/cm^{-1} 3441, 1743, 1548, 1371, 1265, 1055; ^{13}C NMR ($CDCl_3$): δ 164.08 (*C*), 140.33 (*C*), 138.91 (*C*), 138.10 (*C*), 136.91 (*C*), 128.88 (*C*), 128.81 (*C*H), 128.80 (*C*H), 123.42 (*C*, q, *J* 277.0 Hz), 127.52 (*C*H), 126.87 (*C*H), 124.33 (*C*H), 124.24 (*C*), 123.36 (*C*H), 110.61 (*C*H), 101.85 (*C*H), 67.80 (CH_2, q, *J* 35.0 Hz), 61.30 (CH_2), 34.66 (CH_2), 14.28 (CH_3); ^{1}H NMR ($CDCl_3$): δ 8.02 (1H, s, H4), 7.90 (1H, br, H1), 7.59 (1H, dd, *J* 8.6, 1.4 Hz, H6), 7.37–7.34 (2H, m, H4′), 7.30–7.24 (5H, m, H3′, H5′, H1″, H7), 6.37 (1H, d, *J* 1.2 Hz, H3), 4.40–4.30 (4H, m, H3″, H6″), 4.14 (2H, s, H1′), 1.40 (3H, t, *J* 7.0 Hz, H7″); *m*/*z* (EI$^+$) 403 ([M]$^+$, 45), 252 (25), 221 ([M–propanoic acid side chain]$^+$, 100), 91 (27); *m*/*z* HRMS (EI$^+$) [M]$^+$ calcd 403.1395 for $C_{22}H_{20}F_3NO_3$ found 403.1392; Elemental calcd C, 65.5; H, 5.0; N, 3.5; found C, 65.4; H, 4.9; N, 3.5. *E* isomer **31**: pale yellow oil: R_f 0.25 [hexane–EtOAc 7.5 : 2.5]; ν_{max} (DCM film)/cm^{-1} 3441, 1743, 1548, 1265, 1053; ^{13}C NMR ($CDCl_3$): δ 163.53 (*C*), 143.28 (*C*), 138.55 (*C*), 138.26 (*C*), 136.03 (*C*), 128.79 (*C*H), 128.74 (*C*H), 128.54 (C), 123.23 (*C*, q, *J* 277.0 Hz), 126.79 (*C*H), 124.31 (*C*), 122.75 (*C*H), 121.76 (*C*H), 120.95 (*C*H), 110.08 (*C*H), 101.32 (*C*H), 67.64 (CH_2, q, *J* 35.0 Hz), 61.26 (CH_2), 34.65 (CH_2), 13.78 (CH_3); ^{1}H NMR ($CDCl_3$): δ 7.84 (1H, br, H1), 7.53 (1H, s, H4), 7.37–7.33 (2H, m, H3′), 7.30–7.25 (3H, m, H4′, H5′), 7.17 (1H, d, *J* 8.4 Hz, H7), 7.09 (1H, dd, *J* 8.4, 1.6 Hz, H6), 6.71 (1H, s, H1″), 3.61 (1H, br, H3), 4.27 (2H, q, *J* 16.8, 8.4 Hz, H3″), 4.20 (2H, q, *J* 7.2 Hz, H6″), 4.12 (2H, s, H1′), 1.17 (3H, t, *J* 7.0 Hz, H7″); *m*/*z* (EI$^+$) 403 ([M]$^+$, 45), 252 (25), 221 ([M–propanoic acid side chain]$^+$, 100), 91 (27); *m*/*z* HRMS (EI$^+$) [M]$^+$ calcd 403.1395 for $C_{22}H_{20}F_3NO_3$ found 403.1392; Elemental calcd C, 65.5; H, 5.0; N, 3.5; found C, 65.5; H, 5.0; N, 3.4.

Methyl 3-(2-benzyl-1*H*-indol-5-yl)-2-(2,2,2-trifluoroethoxy)-propanoate 37. The reaction was performed according to procedure **B** using **30**. After purification *via* flash column chromatography (hexane–EtOAc 15 : 1, 10 : 1, 9 : 1, 8 : 2) the desired product 48.0 mg (73%) of **37** was obtained as a very pale yellow oil: R_f 0.15 [hexane–EtOAc 8 : 2]; ν_{max} (DCM film)/cm^{-1} 3361, 2920–2846, 1741, 1448, 1275, 1157, 790; ^{13}C NMR ($CDCl_3$): δ 171.50 (*C*), 138.43 (*C*), 138.16 (*C*), 135.37 (*C*), 128.86 (*C*), 128.81 (*C*H), 128.70 (*C*H), 123.74 (*C*, q, *J* 278.0 Hz), 127.28 (*C*), 126.72 (*C*H), 122.85 (*C*H), 120.54 (*C*H), 110.29 (*C*H), 100.88 (*C*H), 82.02 (*C*H), 67.85 (CH_2, q, *J* 34.5 Hz), 52.08 (CH_3), 39.17 (CH_2), 34.69 (CH_2); ^{1}H NMR ($CDCl_3$): δ 7.76 (1H, br, H1), 7.40 (1H, s, H4), 7.36–7.32 (2H, m, H4′), 7.28–7.25 (3H, m, H3′, H5′), 7.17 (1H, d, *J* 8.4 Hz H7), 6.99 (1H, dd, *J* 8.2, 1.4 Hz, H6), 6.28 (1H, br, H3), 4.26–4.23 (1H, m, H2″), 4.13 (2H, s, H1′), 3.99–3.93 (1H, m, H3″), 3.74–3.66 (4H, m, H6″, H3″), 3.21–3.11 (2H, m, H1″); *m*/*z* (EI$^+$) 391 ([M]$^+$, 27), 220 ([M–propanoic acid side chain]$^+$, 100), 49 (23); *m*/*z* HRMS (EI$^+$) [M]$^+$ calcd 391.1395 for $C_{21}H_{20}F_3NO_3$ found 391.1394; Elemental calcd C, 64.4; H, 5.2; N, 3.6; found C, 64.5; H, 5.1; N, 3.5.

3-(2-Benzyl-1*H*-indol-5-yl)-2-(2,2,2-trifluoroethoxy)propanoic acid 5. The synthesis was performed according to procedure **C** using **37**. This gave product **5** 14.0 mg (99%) as a pale yellow oil: R_f 0.20 [hexane–EtOAc 1 : 9]; ν_{max} (DCM film)/cm^{-1} 3289, 2950–2850, 1702, 1651, 1425, 1053; ^{13}C NMR ($CDCl_3$): δ 174.74 (*C*), 138.37 (*C*), 138.31 (*C*), 135.46 (*C*), 128.93 (*C*), 128.83 (*C*H), 128.74 (*C*H), 123.42 (*C*, q, *J* 277.0 Hz), 126.92 (*C*), 126.77 (*C*H), 122.86 (*C*H), 120.67 (*C*H), 110.43 (*C*H), 100.93 (*C*H), 81.63 (*C*H), 68.13 (CH_2, q, *J* 34.5 Hz), 38.98 (CH_2), 34.72 (CH_2); ^{1}H NMR ($CDCl_3$): δ 7.76 (1H, br, H1), 7.43 (1H, s, H4), 7.34–7.32 (2H, m, H4′), 7.28–7.25 (3H, m, H3′, H5′), 7.18 (1H, d, *J* 8.4 Hz, H7), 7.02 (1H, dd, *J* 8.2, 1.4 Hz, H6), 6.29 (1H, d, *J* 0.8 Hz, H3), 4.29 (1H, dd, *J* 8.0, 4.4 Hz, H2″), 4.13 (2H, s, H1′), 3.96–3.87 (1H, m, H3″), 3.77–3.70 (1H, m, H3″), 3.28 (1H, dd, *J* 14.2, 4.2 Hz, H1″), 3.15 (1H, dd, *J* 14.0, 8.0 Hz, H1″); *m*/*z* (EI$^+$) 377 ([M]$^+$, 30), 220 ([M–propanoic acid side chain]$^+$, 100), 142 (9), 91 (12); *m*/*z* HRMS (EI$^+$) [M]$^+$ calcd 377.1239 for $C_{20}H_{18}F_3NO_3$ found 377.1244; Elemental calcd C, 63.7; H, 4.8; N, 3.7; found C, 63.5; H, 4.7; N, 3.7.

2-(Naphthalen-2-ylmethyl)-1-(phenylsulfonyl)-1*H*-indole-5-carbonitrile 23. This reaction was performed by adaptation of a procedure of Iwanowicz *et al.*[67] in a similar way to the synthesis of **22** from **21**, but replacing a solution of benzyl bromide with a solution of 2-(bromomethyl)-naphthalene. After the work up, the oily residue was purified *via* flash column chromatography (hexane–EtOAc 15 : 1) to yield **23** (30%) as yellow oil. Crystallisation from hexane–DCM 20 : 1 afforded shiny white needles: mp 142.0–143.0 °C (hexane–DCM 20 : 1); R_f 0.30 [hexane–EtOAc 7 : 3]; ν_{max} (nujol)/cm^{-1} 2950–2850, 2222, 1651, 1548, 1379, 1055; ^{13}C NMR ($CDCl_3$): δ 143.41 (*C*), 139.07 (*C*), 138.57 (*C*), 134.41 (*C*), 134.09 (*C*H), 133.47 (*C*), 132.44 (*C*), 129.34 (*C*H), 128.36 and 127.95 (*C*H), 127.65, 127.61, 127.45, 127.25, 126.29, 126.25, 125.92, 125.08 (*C*H), 119.32 (*C*), 115.38 (*C*H), 110.30 (*C*H), 107.14 (*C*), 35.24 (CH_2); ^{1}H NMR ($CDCl_3$): δ 8.31 (1H, d, *J* 8.8 Hz, H7), 7.86–7.83 (1H, m), 7.79 (1H, d, *J* 8.4 Hz), 7.73–7.71 (2H, m), 7.63 (2H, dd, *J* 8.6, 1.0 Hz), 7.59 (1H, br, H4), 7.56 (1H, dd, *J* 8.8, 1.6 Hz, H6), 7.50–7.46 (3H, m, H3′), 7.32–7.28 (3H, m), 6.21 (1H, br, H3), 4.53 (2H, s, H5′); *m*/*z* (EI$^+$) 422 ([M]$^+$, 61), 280 ([M–methyl naphthyl group]$^+$, 100), 141 (34), 77 (34); *m*/*z* HRMS (EI$^+$) [M]$^+$

calcd 422.1089 for $C_{26}H_{18}N_2O_2S$ found 422.1087; Elemental calcd C, 73.9; H, 4.3; N, 6.6; found C, 74.0; H, 4.2; N, 6.6. *Crystal data for* **23**: $C_{26}H_{18}N_2O_2S$, $M = 422.48$, monoclinic, $P2_1/n$ (no. 14), $a = 18.0309(4)$, $b = 5.88641(15)$, $c = 19.5220(4)$ Å, $\beta = 98.636(2)°$, $V = 2048.52(19)$ Å^3, $Z = 4$, $D_c = 1.370$ g cm^{-3}, μ(Cu-Kα) = 1.615 mm^{-1}, $T = 173$ K, colourless needles, Oxford Diffraction Xcalibur PX Ultra diffractometer; 3230 independent measured reflections, F^2 refinement, $R_1 = 0.040$, w$R_2 = 0.076$, 1894 independent observed absorption-corrected reflections [$|F_o| > 4\sigma(|F_o|)$, $2\theta_{max} = 126°$], 280 parameters. CCDC 667595.

2-(Naphthalen-2-ylmethyl)-1*H*-indole-5-carbonitrile 25. The reaction was carried out according to procedure **E** using starting material **23**. This procedure gave 335 mg (99%) of **25** as an amorphous white powder: mp 152.0–153.0 °C (hexane–DCM 20 : 1); R_f 0.25 [hexane–EtOAc 7 : 3]; ν_{max} (nujol)/cm^{-1} 3206, 2980–2850, 2215, 1461, 1371, 1053; ^{13}C NMR ($CDCl_3$): δ 140.25 (*C*), 137.95 (*C*), 134.96 (*C*), 133.54 (*C*), 132.47 (*C*), 128.76 (*C*H), 128.51 (*C*), 127.74, 127.54, 127.30, 127.23, 126.99, 126.47, 125.99, 125.44, 124.53 (*C*H), 120.81 (*C*), 111.28 (*C*H), 102.88 (*C*), 101.77 (*C*H), 34.79 (CH_2); ^{1}H NMR ($CDCl_3$): δ 8.13 (1H, br, H1), 7.89 (1H, s), 7.86–7.80 (3H, m), 7.73 (1H, s, H4), 7.54–7.47 (2H, m), 7.37–7.34 (2H, m, H6), 7.27 (1H, d, *J* 8.4 Hz, H7), 6.46 (1H, d, *J* 0.8 Hz, H3), 4.33 (2H, s, H1′); *m*/*z* (EI$^+$) 282 ([M]$^+$, 100), 155 (52); *m*/*z* HRMS (EI$^+$) [M]$^+$ calcd 282.1157 for $C_{20}H_{14}N_2$ found 282.1142; Elemental calcd C, 85.1; H, 5.0; N, 9.9; found C, 85.1; H, 4.9; N, 9.9.

2-(Naphthalen-2-ylmethyl)-1*H*-indole-5-carbaldehyde 27. The reaction was carried out according to procedure **F** using **25**. After flash column chromatography (hexane–EtOAc 12 : 1, 10 : 1, 9 : 1, 7 : 3), 272 mg (86%) of **27** was obtained as an amorphous pale yellow powder: mp 162.5–163.5 °C (hexane–DCM 20 : 1); R_f 0.20 [hexane–EtOAc 7.5 : 2.5]; ν_{max} (nujol)/cm^{-1} 3221, 2980–2850, 1660, 1306, 796; ^{13}C NMR ($CDCl_3$/MeOD 1 : 2): δ 194.21 (*C*H), 141.47 (*C*), 141.20 (*C*), 136.49 (*C*), 134.11 (*C*), 132.84 (*C*), 129.29 (*C*), 129.08 (*C*), 128.65, 128.03, 127.96, 127.63, 127.51, 126.54, 126.02 (*C*H), 125.93 (*C*H), 121.94 (*C*H), 111.83 (*C*H), 102.50 (*C*H), 35.13 (CH_2); ^{1}H NMR ($CDCl_3$/MeOD 1 : 2) δ 9.86 (1H, s, H1″), 8.00 (1H, s, H4), 7.78–7.73 (3H, m), 7.69 (1H, s, H1), 7.60 (1H, d, *J* 8.4 Hz, H6), 7.44–7.33 (4H, m, H7), 6.36 (1H, br, H3), 4.25 (2H, s, H1′); *m*/*z* (EI$^+$) 285 ([M]$^+$, 100), 254 ([M–aldehyde]$^+$, 38), 158 (50), 158 (42); *m*/*z* HRMS (EI$^+$) [M]$^+$ calcd 285.1154 for $C_{20}H_{15}NO$ found 285.1146; Elemental calcd C, 84.2; H, 5.3; N, 4.9; found C, 84.1; H, 5.3; N, 4.9.

(*Z*)-Ethyl 2-ethoxy-3-(2-(naphthalen-2-ylmethyl)-1*H*-indol-5-yl)-acrylate 32 and (*E*)-ethyl 2-ethoxy-3-(2-(naphthalen-2-ylmethyl)-1*H*-indol-5-yl)acrylate 33. The synthesis was conducted according to procedure **A** using the staring materials **27** and **9**. After flash column chromatography (hexane–EtOAc 12 : 1, 9 : 1, 8 : 2, EtOAc) 112 mg (79%) of two isomers was obtained. Ratio $Z : E = 63 : 37$. *Z* isomer **32**: colourless oil: R_f 0.30 [hexane–EtOAc 7 : 3]; ν_{max} (DCM film)/cm^{-1} 3352, 3053–2900, 1701, 1620, 1252, 1095, 740; ^{13}C NMR ($CDCl_3$): δ 165.32 (*C*), 142.39, 138.60, 136.60, 135.70, 133.52, 132.36, 128.81 (*C*), 128.45, 127.67, 127.53, 127.16, 127.12, 126.34, 126.26, 125.74 (*C*H), 125.48 (*C*), 124.07 (*C*H), 122.70 (*C*H), 110.50 (*C*H), 101.81 (*C*H), 67.40 (CH_2), 60.87 (CH_2), 34.83 (CH_2), 15.54 (CH_3), 14.32 (CH_3); ^{1}H NMR ($CDCl_3$): δ 8.06 (1H, s, H4), 7.98 (1H, br, H1), 7.86–7.78 (3H, m), 7.72 (1H, s), 7.65 (1H, dd, *J* 8.6, 1.0 Hz, H6), 7.52–7.47 (2H, m), 7.38 (1H, dd, *J* 8.4, 1.2 Hz), 7.23–7.21 (1H, d, *J* 8.4, H7), 7.19 (1H, s), 6.41 (1H, br, H3), 4.34–4.29 (4H, m, H6″, H1′), 4.03 (2H, q, *J* 7.0 Hz, H3″), 1.43–1.37 (6H, m, H4″, H7″); *m*/*z* (EI$^+$) 399 ([M]$^+$, 78), 342 ([M–both ethyl groups]$^+$, 25), 297 (24), 141 (100); *m*/*z* HRMS (EI$^+$) [M]$^+$ calcd 399.1834 for $C_{26}H_{25}NO_3$ found 399.1830; Elemental calcd C, 78.2; H, 6.3; N, 3.5; found C, 78.3; H, 6.2; N, 3.6. *E* isomer **33**: colourless oil: R_f 0.30 [hexane–EtOAc 7 : 3]; ν_{max} (DCM film)/cm^{-1} 3392, 3050–2900, 1712, 1635, 1228, 1153, 740; ^{13}C NMR ($CDCl_3$): δ 165.19 (*C*), 146.09, 138.06, 135.90, 135.46, 133.52, 132.33, 128.67 (*C*), 128.39, 127.66, 127.52, 127.19, 127.07, 126.23 (*C*H), 126.20 (*C*), 125.69 (*C*H), 122.37 (*C*H), 119.92 (*C*H), 111.61 (*C*H), 110.05 (*C*H), 101.25 (*C*H), 64.60 (CH_2), 61.04 (CH_2), 34.70 (CH_2), 14.53 (CH_3), 13.73 (CH_3); ^{1}H NMR ($CDCl_3$): δ 7.85–7.79 (4H, m), 7.71 (1H, s), 7.52–7.45 (2H, m), 7.42 (1H, s, H4), 7.37 (1H, dd, *J* 8.4, 1.6 Hz), 7.13 (1H, d, *J* 8.4 Hz, H7), 6.99 (1H, dd, *J* 8.4, 1.6 Hz, H6), 6.33 (1H, d, *J* 0.8 Hz, H3), 6.29 (1H, s, H1″), 4.28 (2H, s, H1′), 4.15 (2H, q, *J* 7.2 Hz, H6″), 3.95 (2H, q, *J* 13.8, 7.0 Hz, H3″), 1.42 (3H, t, *J* 6.8 Hz, H4″), 1.09 (3H, t, H7″); *m*/*z* (EI$^+$) 399 ([M]$^+$, 72), 342 ([M–both ethyl groups]$^+$, 24), 282 (34), 141 (100); *m*/*z* HRMS (EI$^+$) [M]$^+$ calcd 399.1834 for $C_{26}H_{25}NO_3$ found 399.1832; Elemental calcd C, 78.2; H, 6.3; N, 3.5; found C, 78.3; H, 6.2; N, 3.6.

Methyl 2-ethoxy-3-(2-(naphthalen-2-ylmethyl)-1*H*-indol-5-yl)-propanoate 38. The reaction was performed according to procedure **B** using **32**. After purification *via* flash column chromatography (hexane–EtOAc 10 : 1, 9 : 1, 8 : 2) 85.0 mg (80%) of **38** was obtained as a light yellow oil: R_f 0.15 [hexane–EtOAc 8 : 2]; ν_{max} (DCM film)/cm^{-1} 3392, 3050–2900, 1741, 1643, 1273, 1114, 740; ^{13}C NMR ($CDCl_3$): δ 173.30 (*C*), 137.89, 135.97, 135.30, 133.53, 133.18, 132.33, 128.90 (*C*), 128.41 (*C*H), 128.26 (*C*), 127.72, 127.52, 127.24, 127.09, 126.35, 125.69 (*C*H), 122.95 (*C*H), 120.47 (*C*H), 110.19 (*C*H), 101.02 (*C*H), 81.02 (*C*H), 66.20 (CH_2), 51.77 (CH_3), 39.53 (CH_2), 34.92 (CH_2), 15.07 (CH_3); ^{1}H NMR ($CDCl_3$): δ 7.84–7.79 (4H, m), 7.71 (1H, s, H1), 7.51–7.45 (2H, m), 7.42 (1H, s, H4), 7.38 (1H, dd, *J* 8.4, 1.6 Hz), 7.14 (1H, d, *J* 8.4 Hz, H7), 7.00 (1H, dd, *J* 8.0, 1.6 Hz, H6), 6.33 (1H, d, *J* 1.2 Hz, H3), 4.28 (2H, s, H1′), 4.09 (1H, dd, *J* 7.2, 6.0 Hz, H2″), 3.71 (3H, s, H6″), 3.63–3.58 (1H, m, H3″), 3.40–3.33 (1H, m, H3″), 3.11 (1H, s, H1″), 3.09 (1H, d, *J* 1.6 Hz, H1″), 1.17 (3H, t, *J* 6.8 Hz, H4″); *m*/*z* (EI$^+$) 387 ([M]$^+$, 27), 270 ([M–side chain on 5-position]$^+$, 100), 146 (67); *m*/*z* HRMS (EI$^+$) [M]$^+$ calcd 387.1834 for $C_{25}H_{25}NO_3$ found 387.1835; Elemental calcd C, 77.5; H, 6.5; N, 3.6; found C, 77.5; H, 6.4; N, 3.7.

2-Ethoxy-3-(2-(naphthalen-2-ylmethyl)-1*H*-indol-5-yl)propanoic acid 6. The synthesis was performed according to procedure **C** using **38**. This procedure resulted in 58.5 mg (98%) of **6** as a yellow oil: R_f 0.25 [hexane–EtOAc 1 : 9]; ν_{max} (DCM film)/cm^{-1} 3403, 3053–2900, 1720, 1644, 1267, 1110, 739; ^{13}C NMR ($CDCl_3$): δ 174.59 (*C*), 138.03, 135.90, 135.39, 133.53, 132.35, 128.87 (*C*), 128.44 (*C*H), 127.68 (*C*H), 127.65 (*C*), 127.53, 127.24, 127.12, 126.25, 125.72 (*C*H), 123.01 (*C*H), 120.69 (*C*H), 110.32 (*C*H), 101.04 (*C*H), 80.31 (*C*H), 66.89 (CH_2), 38.75 (CH_2), 34.93 (CH_2), 15.07 (CH_3); ^{1}H NMR ($CDCl_3$): δ 7.85–7.79 (4H, m), 7.71 (1H, s), 7.51–7.45 (2H, m), 7.43 (1H, s, H4), 7.38 (1H, dd, *J* 8.4, 1.6 Hz), 7.15 (1H, d, *J* 8.4 Hz, H7), 7.01 (1H, dd, *J* 8.2, 1.4 Hz, H6), 6.34 (1H, br, H3), 4.28 (2H, s, H1′), 4.12 (1H, dd, *J* 7.6, 4.4 Hz, H2″), 3.62–3.54 (1H, m, H3″), 3.50–3.42 (1H, m, H3″), 3.23 (1H, dd, *J*

14.2, 4.2 Hz, H1″), 3.08 (1H, dd, *J* 14.2, 7.8 Hz, H1″), 1.18 (3H, t, *J* 7.0 Hz, H4″); *m/z* (EI$^+$) 373 ([M]$^+$, 33), 270 ([M−side chain on 5-position]$^+$, 100), 142 (36); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 373.1678 for $C_{24}H_{23}NO_3$ found 373.1676; Elemental calcd C, 77.2; H, 6.2; N, 3.8; found C, 73.2; H, 6.2; N, 3.0.

(*Z*)-Ethyl 3-(2-(naphthalen-2-ylmethyl)-1*H*-indol-5-yl)-2-(2,2,2-trifluoroethoxy)-acrylate 34 and (*E*)-ethyl 3-(2-(naphthalen-2-ylmethyl)-1*H*-indol-5-yl)-2-(2,2,2-trifluoro-ethoxy)-acrylate 35. The synthesis was carried out according to procedure **D** using the starting materials **27** and **10**. After purification *via* flash column chromatography (hexane–EtOAc 15 : 1, 10 : 1), 183 mg (78%) of two isomers was obtained. Ratio *Z* : *E* = 38 : 62. *Z* isomer **34**: almost colourless oil: R_f 0.35 [hexane–EtOAc 7 : 3]; ν_{max} (DCM film)/cm^{-1} 3408, 2975, 1702, 1620, 1263, 1167, 742; ^{13}C NMR ($CDCl_3$): δ 164.08 (*C*), 140.38, 138.80, 136.95, 135.56, 133.56, 132.42, 128.93 (*C*), 128.58 (*C*H), 127.71 (*C*H), 123.43 (*C*, q, *J* 277.0 Hz), 127.55, 127.48, 127.18, 127.15 126.34, 125.82 (*C*H), 124.38 (*C*H), 124.30 (*C*), 123.39 (*C*H), 110.64 (*C*H), 102.00 (*C*H), 67.82 (*C*H$_2$, q, *J* 35.0 Hz), 61.30 (*C*H$_2$), 34.88 (*C*H$_2$), 14.28 (*C*H$_3$); ^{1}H NMR ($CDCl_3$): δ 8.06 (1H, s, H4), 7.93 (1H, br, H1), 7.88–7.82 (3H, m), 7.75 (1H, s), 7.61 (1H, dd, *J* 8.6, 1.4 Hz, H6), 7.56–7.48 (2H, m), 7.43 (1H, dd, *J* 8.4, 1.6 Hz), 7.29 (1H, s, H1″), 7.24 (1H, d, *J* 8.4 Hz, H7), 6.45 (1H, d, *J* 0.8 Hz, H3), 4.43 (4H, m, H3″, H6″), 4.32 (2H, s, H1′), 1.41 (3H, t, *J* 7.2 Hz, H7″); *m/z* (EI$^+$) 453 ([M]$^+$, 100), 342 (17), 141 (93); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 453.1552 for $C_{26}H_{22}F_3NO_3$ found 453.1553; Elemental calcd C, 68.9; H, 4.9; N, 3.1; found C, 68.8; H, 4.8; N, 2.9. *E* isomer **35**: off-white powder: mp 123.0–124.5 °C (hexane–DCM 20 : 1); R_f 0.30 [hexane–EtOAc 7 : 3]; ν_{max} (DCM film)/cm^{-1} 3441, 2970–2850, 1709, 1645, 1267, 1132, 740; ^{13}C NMR ($CDCl_3$): δ 163.51 (*C*), 143.33, 138.42, 136.08, 135.73, 133.54, 132.38, 128.58 (*C*), 128.49, 127.70, 127.53 (*C*H), 123.25 (*C*H$_2$, q, *J* 277.0 Hz), 127.16, 127.13, 126.31, 125.78 (*C*H), 124.37 (*C*), 122.82 (*C*H), 121.82 (*C*H), 121.00 (*C*H), 110.11 (*C*H), 101.49 (*C*H), 67.68 (*C*H$_2$, q, *J* 35.0 Hz), 61.26 (*C*H$_2$), 34.87 (*C*H$_2$), 13.79 (*C*H$_3$); ^{1}H NMR ($CDCl_3$): δ 7.86–7.79 (4H, m), 7.71 (1H, s, H1), 7.55 (1H, s, H4), 7.53–7.49 (2H, m), 7.36 (1H, dd, *J* 8.4, 1.6 Hz), 7.16 (1H, d, *J* 8.4 Hz, H7), 7.09 (1H, dd, *J* 8.4, 1.6 Hz), 6.72 (1H, s, H1″), 6.36 (1H, d, *J* 0.8 Hz, H3), 4.30–4.23 (4H, m, 2xH3″, 2xH1′), 4.20 (2H, q, *J* 7.2 Hz, H6″), 1.17 (3H, t, *J* 7.2 Hz, H7″); *m/z* (EI$^+$) 453 ([M]$^+$, 61), 342 (10), 210 (100), 141 (46), 105 (69); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 453.1552 for $C_{26}H_{22}F_3NO_3$ found 453.1548; Elemental calcd C, 68.9; H, 4.9; N, 3.1; found C, 69.0; H, 5.0; N, 3.0.

Methyl 3-(2-(naphthalen-2-ylmethyl)-1*H*-indol-5-yl)-2-(2,2,2-trifluoroethoxy)-propanoate 39. The reaction was performed according to the procedure **B** using **34**. After purification *via* flash column chromatography (hexane–EtOAc 9 : 1, 8 : 2) 11.3 mg (79%) of **39** was obtained as a colourless oil: R_f 0.25 [hexane–EtOAc 7.5 : 2.5]; ν_{max} (DCM film)/cm^{-1} 3403, 3054–2850, 1743, 1645, 1279, 1165, 739; ^{13}C NMR ($CDCl_3$): δ 171.49 (*C*), 138.05, 135.90, 135.44, 133.56, 132.38, 128.92 (*C*), 128.46 (*C*H), 123.58 (*C*, q, *J* 277.0 Hz), 127.69 (*C*H), 127.55 (*C*H), 127.36 (*C*), 127.24, 127.13, 126.26, 125.73 (*C*H), 122.93 (*C*H), 120.60 (*C*H), 110.33 (*C*H), 101.07 (*C*H), 82.04 (*C*H), 67.89 (*C*H$_2$, q, *J* 34.5 Hz), 52.09 (*C*H$_3$), 39.18 (*C*H$_2$), 34.94 (*C*H$_2$); ^{1}H NMR ($CDCl_3$): δ 7.85–7.79 (3H, m), 7.77 (1H, br, H1), 7.72 (1H, s), 7.52–7.46 (2H, m), 7.42 (1H, s, H4), 7.38 (1H, dd, *J* 8.6, 1.8 Hz), 7.16 (1H, d, *J* 8.0 Hz, H7), 6.99 (1H, dd, *J* 8.4, 1.6 Hz, H6), 6.34 (1H, m, H3), 4.29 (2H, s, H1′), 4.25 (1H, dd, *J* 7.6, 4.8 Hz, H2″), 4.01–3.92 (1H, m, H3″), 3.74 (3H, s, H6″), 3.74–3.66 (1H, m, H3″), 3.23–3.18 (1H, m, H1″), 3.17–3.12 (1H, m, H1″); *m/z* (EI$^+$) 441 ([M]$^+$, 56), 270 ([M−propanoate side chain]$^+$, 100); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 441.1552 for $C_{25}H_{22}F_3NO_3$ found 441.1548; Elemental calcd C, 68.2; H, 5.0; N, 3.2; found C, 72.7; H, 4.9; N, 3.8.

***3*-(2-(Naphthalen-2-ylmethyl)-1*H*-indol-5-yl)-2-(2,2,2-trifluoro-ethoxy)propanoic acid 7.** The synthesis was performed according to procedure **C** using **39**, giving 5.60 mg (98%) of **7** as a pale yellow oil: R_f 0.40 [Solvent A]; ν_{max} (DCM film)/cm^{-1} 3441, 3000–2700, 1706, 1646, 1267, 1049, 740; δ_C ($CDCl_3$) 174.21 (*C*), 138.19, 135.85, 135.50, 133.55, 132.38, 128.96 (*C*), 128.48, 127.69 (*C*H), 123.43 (*C*, q, *J* 277.0 Hz), 127.54, 127.23, 127.14 (3x*C*H), 126.98 (*C*), 126.27, 125.74 (*C*H), 122.91 (*C*H), 120.70 (*C*H), 110.46 (*C*H), 101.07 (*C*H), 81.65 (*C*H), 68.13 (*C*H$_2$, q, *J* 34.5 Hz), 38.98 (*C*H$_2$), 34.93 (*C*H$_2$); ^{1}H NMR ($CDCl_3$/MeOD 2 : 1): δ 9.53 (1H, br, H6″), 7.76–7.71 (3H, m), 7.65 (1H, s), 7.51 (1H, s, H4), 7.42–7.38 (2H, m), 7.34 (1H, dd, *J* 8.4, 1.2 Hz), 7.16 (1H, d, *J* 8.4 Hz, H7), 7.09–7.07 (1H, m, H6), 6.59 (1H, s, H1″), 6.18 (1H, br, H3), 4.26–4.20 (4H, m, H3″, H1′); *m/z* (EI$^+$) 427 ([M]$^+$, 47), 270 (100), 141 (53); *m/z* HRMS (EI$^+$) [M]$^+$ calcd 427.1395 for $C_{24}H_{20}F_3NO_3$ found 427.1389; Elemental calcd C, 67.4; H, 4.7; N, 3.3; found C, 67.3; H, 4.6; N, 3.2.

PPAR activation assay

A plasmid encoding for each subtype of PPARs, as well as the luciferase reporter gene was transfected into human breast cancer MCF-7 cells. The firefly luciferase catalyses the bioluminescent oxidation of luciferin in the presence of ATP, magnesium and oxygen.[68] The amount of luciferase protein expressed was measured on a luminometer and expressed in relative light units (RLU), used to quantify the efficiency of the PPAR activation.

The indole compounds **2** to **7** were dissolved in DMSO, 0.1% (v/v). DMSO concentration was kept constant in all samples. Human MCF-7 breast cancer cells were seeded at a density of 85,000 cells per well of a 12-well plate. The following day, they were transiently transfected by the SuperFect transfection procedure according to the manufacturer's protocol (Qiagen), using 0.9 μg PPREx3-LUC reporter plasmid and 0.15 μg of the plasmid vector, pcDNA3.1 humanPPARα/γ/δ.

The total amount of plasmid was kept constant at 2.65 μg by compensating with empty vector pCMV-5. Post 24 h after transfection, the cells were treated with the compounds to be tested or the PPAR specific agonists. 48 h after transfection, the cells were washed once with PBS, lysed in 80 μL lysis buffer containing 25 mM TAE, 2 mM DTT, 10% (v/v) glycerol, 1% (v/v) Triton X-100 and de-ionised water. 35 μL of the cell extracts were used for luciferase determination on a LUCY-1 luminometer (Anthos, Austria). The luciferase assay was performed in accordance with the protocol of the Luciferase Assay Kit (BIO Thema AB, Sweden).

Statistical Analysis

Statistical analyses were carried out using GraphPad Prism version 5.0a. All *P* values were two-sided and values <0.05 were considered statistically significant. Values are given as means with standard deviations (SD). One-way analysis of variance

and Tukey's multiple comparison test were used to demonstrate significant differences.

Conclusion

A novel series of 2,5-disubstituted indoles, designed using a software program PROTOBUILD, have been synthesised successfully and shown to be dual PPARα/γ agonists using a luciferase reporter gene activation assay. Further tests are now required in order to assess *in vivo* efficacy and ADME-toxicology properties in order to allow selection of the best candidate for future development. The data reported here demonstrate the potential utility of a new software program PROTOBUILD to predict alternative, novel drug leads for a target that has been the subject of intense interest. Our next step will be to explore the design, synthesis and evaluation of potential drug leads based upon completely novel chemotypes generated by our software program PROTOBUILD.

Abbreviations

EI	electron ionisation
PPAR	peroxisome proliferator-activated receptor
PPRE	peroxisome proliferator response element
LBD	ligand binding domain
VLDL	very low-density lipoprotein

Acknowledgements

We wish to thank Mitsubishi Chemical Corporation, IC-Vec Ltd, and the EPSRC for their financial support of the Imperial College Genetic Therapies Centre. The authors also wish to acknowledge Dr Andrew White for the X-ray crystallography structures, John Barton for high resolution mass spectrometry and Stephen Boyer for elemental analysis.

References

1 A. M. Rouhi, Metabolic Syndrome, *C & E News*, 2004, 83.
2 C. L. Scott, *Am. J. Cardiol.*, 2003, **92**, 35i.
3 M. P. Stern, K. Williams, C. González-Villalpando, K. J. Hunt and S. M. Haffner, *Diabetes Care*, 2004, **27**, 2676.
4 I. Issemann and S. Green, *Nature*, 1990, **347**, 645.
5 C. Dreyer, G. Krey, H. Keller, F. Givel and G. Helftenbein, *Cell*, 1992, **68**, 879.
6 T. M. Willson, P. J. Brown, D. D. Sternbach and B. R. Henke, *J. Med. Chem.*, 2000, **43**, 527.
7 B. Desvergne and W. Wahli, *Endocr. Rev.*, 1999, **20**, 649.
8 W. Wahli, O. Braissant and B. Desvergne, *Chem. Biol.*, 1995, **2**, 261.
9 J. Berger and D. E. Moller, *Annu. Rev. Med.*, 2002, **53**, 409.
10 A. K. Hihi, L. Michalik and W. Wahli, *Cell. Mol. Life Sci.*, 2002, **59**, 790.
11 T. Lemberger, B. Desvergne and W. Wahli, *Annu. Rev. Cell Dev. Biol.*, 1996, **12**, 335.
12 G. D. Barish, V. A. Narkar and R. M. Evans, *J. Clin. Invest.*, 2006, **116**, 590.
13 P. Lefebvre, G. Chinetti, J. C. Fruchart and B. Staels, *J. Clin. Invest.*, 2006, **116**, 571.
14 R. K. Semple, V. K. K. Chatterjee and S. O'Rahilly, *J. Clin. Invest.*, 2006, **116**, 581.
15 K. Schoonjans, B. Staels and J. Auwerx, *Biochim. Biophys. Acta.*, 1996, **1302**, 93.
16 O. Braissant, F. Foufelle, C. Scotto, M. Dauca and W. Wahli, *Endocrinology*, 1996, **137**, 354.
17 K. Schoonjans, M. Watanabe, H. Suzuki, A. Mahfoudi, G. Krey, W. Wahli, P. Grimaldi, B. Staels, T. Yamamoto and J. Auwerx, *J. Biol. Chem.*, 1995, **270**, 19269.
18 S. M. Rangwala and M. A. Lazar, *TIPS*, 2004, **25**, 331.
19 P. Kirkpatrick, *Nat. Rev. Drug Discovery*, 2002, **1**, 169.
20 M. W. Schwartz and S. E. Kahn, *Nature*, 1999, **402**, 860.
21 G. Martin, K. Schoonjans, A. M. Lefebvre, B. Staels and J. Auwerx, *J. Biol. Chem.*, 1997, **272**, 28210.
22 W. Wahli, *Swiss Medical Weekly*, 2002, **132**, 83.
23 W. R. Oliver, J. L. Shenk, M. R. Snaith, C. S. Russell, K. D. Plunket, N. L. Bodkin, M. C. Lewis, D. A. Winegar, M. L. Sznaidman, M. H. Lambert, H. E. Xu, D. D. Sternbach, S. A. Kliewer, B. C. Hansen and T. M. Willson, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 5306.
24 B. Staels, J. Dallongeville, J. Auwerx, K. Schoonjans, E. Leitersdorf and J. C. Fruchart, *Circulation*, 1998, **98**, 2088.
25 P. Raskin, E. B Rappaport, S. T Cole, Y. Yan, R. Patwardhan and M. I. Freed, *Diabetologia*, 2000, **43**, 278.
26 J. Berger, P. Bailey, C. Biswas, C. A. Cullinan, T. W. Doebber, N. S. Hayes, R. Saperstein, R. G. Smith and M. D. Leibowitz, *Endocrinology*, 1996, **137**, 4189.
27 A. R. Miller, *Drug Dev. Res.*, 2006, **67**, 574.
28 B. Gross and B. Staels, *Best Pract. Res. Clin. Endocrinol. Metab.*, 2007, **21**, 687.
29 A.-M. Lefebvre, J. Peinado-Onsurbe, I. Leitersdorf, M. R. Briggs, J. R. Paterniti, J. C. Fruchart, C. Fievet, J. Auwerx and B. Staels, *Arterioscler. Thromb. Vasc. Biol.*, 1997, **17**, 1756.
30 P. Gervois, J. C. Fruchart and B. Staels, *Int. J. Clin. Pract.*, 2004, **58**, 22.
31 H. J. Bohm, *J. Comput.-Aided Mol. Des.*, 1992, **6**, 61.
32 M. Stahl, N. P. Todorov, T. James, H. Mauser, H.-J. Bohm and P. M. Dean, *J. Comput.-Aided Mol. Des.*, 2002, **16**, 459.
33 B. A. Grzybowski, A. V. Ishchenko, C. Y. Kim, G. Topalov, R. Chapman, D. W. Christianson, G. M. Whitesides and E. I. Shakhnovich, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 1270.
34 R. S. Bohacek and C. McMartin, *J. Am. Chem. Soc.*, 1994, **116**, 5560.
35 The Molecular Operating Environment (MOE) available under license from Chemical Computing Group Inc., 1010 Sherbrooke St. W. Suite 910, Montreal, Quebec, Canada H3A 2R7.
36 A. R. Organov and C. W. Glass, *J. Chem. Phys.*, 2006, **124**, 244704.
37 H. M. Vinkers, M. R. de Jonge, F. F. Daeyaert, J. Heeres, L. M. Koymans, J. H. van Lenthe, P. J. Lewi, H. Timmerman, K. Van Aken and P. A. Janssen, *J. Med. Chem.*, 2003, **46**, 2765.
38 M. Clark, R. D. Cramer and N. Van Opdenbosch, *J. Comput. Chem.*, 1989, **10**, 982.
39 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 490.
40 P. A. Nielsen and T. Liljefors, *J. Comput.-Aided Mol. Des.*, 2001, **15**, 753.
41 H-J. Bohm, *J. CAMD.*, 1998, **12**, 309.
42 F. J Carver, C. A. Hunter and E. M. Seward, *Chem. Commun.*, 1998, 775.
43 C. A. Hunter, J. Singh and J. M. Thornton, *J. Mol. Biol.*, 1991, **218**, 837.
44 M. L. Verdonk, G. J. Boks, H. Kooijman, J. A. Kanters and J. Kroon, *J. Comput.-Aided Mol. Des.*, 1993, **7**, 173.
45 R. J. Zauhar, C. L. Colbert, R. S. Morgan and W. J. Welsh, *Biopolymers*, 2000, **53**(3), 233.
46 J. E. Cobb, B. R. Henke and S. G. Blanchard, *Diabetologia*, 2000, **43**, A189.
47 J. Xu and J. Stevenson, *J. Chem. Inf. Comp. Sci.*, 2000, **40**, 1177.
48 C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Delivery Rev.*, 1997, **23**, 3.
49 X. Q. Lewell, D. B. Judd, S. P. Watson and M. M. Rhann, *J. Chem. Inf. Comp. Sci.*, 1998, **38**, 511.
50 J. Rudolph, L. Chen, D. Majumdar, W. H. Bullock, M. Burns, T. Claus, F. E. Dela Cruz, M. Daly, F. J. Ehrgott, J. S. Johnson, J. N. Livingston, R. W. Schoenleber, J. Shapiro, L. Yang, M. Tsutsumib and X. Ma, *J. Med. Chem.*, 2007, **50**, 984.
51 B. Kuhn, H. Hilpert, J. Benz, A. Binggeli, U. Grether, R. Humm, H. P. Marki, M. Meyer and P. Mohr, *Bioorg. Med. Chem. Lett.*, 2006, **16**, 4016.
52 N. Mahindroo, C. F. Huang, Y. H. Peng, C. C. Wang, C. C. Liao, T. W. Lien, S. K. Chittimalla, W. J. Huang, C. H. Chai, E. Prakash, C. P. Chen, T. A. Hsu, C. H. Peng, I. L. Lu, L. H. Lee, Y. W. Chang, W.

C. Chen, Y. C. Chou, C. T. Chen, C. M. Goparaju, Y. S. Chen, S. J. Lan, M. C. Yu, X. Chen, Y. S. Chao, S. Y. H. Wu and P. Hsieh, *J. Med. Chem.*, 2005, **48**, 8194.
53 N. Mahindroo, C. C. Wang, C. C. Liao, C. F. Huang, I. L. Lu, T. W. Lien, Y. H. Peng, W. J. Huang, Y. T. Lin, M. C. Hsu, C. H. Lin, C. H. Tsai, J. T. Hsu, X. Chen, P. C. Lyu, Y. S. Chen, S. Y. Wu and H. P. Hsieh, *J. Med. Chem.*, 2006, **49**, 1212.
54 N. Mahindroo, Y. H. Peng, C. H. Lin, U. K. Tan, E. Prakash, T. W. Lien, I. L. Lu, H. J. Lee, J. T. Hsu, X. Chen, C. C. Liao, P. C. Lyu, Y. S. Chao, S. Y. Wu and H. P. Hsieh, *J. Med. Chem.*, 2006, **49**, 6421.
55 B. B. Lohray, V. B. Lohray, A. C. Bajji, S. Kalchar and R. Ramanujam, *US patent*, 2000, **6**(054), 453.
56 W. Zhang, X. Y. Cao, H. Zi and J. Pei, *Org. Lett.*, 2005, **7**, 959.
57 W. Grell and H. Machleidt, *Annal. Chem.*, 1966, **699**, 53.
58 L. McElweewhite, W. A Goddard and D. A. Dougherty, *J. Am. Chem. Soc.*, 1984, **106**, 3461.
59 C. J. Moody, E. Sie and J. J. Kulagowski, *Tetrahedron*, 1992, **48**, 3991.
60 M. Regitz, W. Anschutz and A. Liedhegener, *Chem. Ber.*, 1968, **101**, 3734.
61 D. Haigh, H. C. Birrell, B. C. C. Cantello, R. M. Hindley, A. Ramaswamy, H. K. Rami and N. C. Stevens, *Tetrahedron: Asymmetry*, 1999, **10**, 1335.
62 R. J. Sundberg and H. F. Russell, *J. Org. Chem.*, 1973, **38**, 3324.
63 R. T. Nolte, G. B Wisely, S. Westin, J. E. Cobb, M. H. Lambert, R. Kurokawa, M. G. Rosenfeld, T. M. Willson, C. K. Glass and M. V. Milburn, *Nature*, 1998, **395**, 137.
64 H. E. Xu, M. H. Lambert, V. G Montana, K. D. Plunket, L. B. Moore, J. L. Collins, J. A. Oplinger, S. A. Kliewer, R. T. Gampe Jr, D. D. McKee, J. T. Moore and T. M. Willson, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 13919.
65 W. C. Still, M. Kahn and A. Mitra, *J. Org. Chem.*, 1978, **43**, 2923.
66 D. Haigh, *Tetrahedron*, 1994, **50**, 3177.
67 E. J. Iwanowicz, W. F. Lau, J. Lin, D. G. M. Roberts and S. M. Seiler, *Bioorg. Med. Chem. Lett.*, 1996, **6**, 1339.
68 J. R. de Wet, K. V. Wood, M. Deluca, D. R. Helinski and S. Subramani, *Mol. Cell. Biol.*, 1987, **7**, 725.

