Bioorganic & Medicinal Chemistry Letters 20 (2010) 6616–6619

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

Identification of optimum computational protocols for modeling the aryl hydrocarbon receptor (AHR) and its interaction with ligands

Ashutosh S. Jogalekar *, Stephan Reiling, Roy J. Vaz

Structure, Design, Informatics, sanofi-aventis US, 1041 Rt. 202-206, Bridgewater, NJ 08807, United States

ARTICLE INFO

Article history:
Received 15 June 2010
Revised 2 September 2010
Accepted 7 September 2010
Available online 15 September 2010

Keywords:
Aryl hydrocarbon receptor (AHR)
Docking
Scoring
Molecular dynamics
MM-GBSA
CYP1A1
TCDD

ABSTRACT

The aryl hydrocarbon receptor (AHR) is one of the principal xenobiotic receptors in living organisms and is responsible for interacting with several drugs and environmental toxins, most notably tetrachlorodibenzodioxin (TCDD). Binding of diverse agonists to AHR initiates an extensive set of downstream gene expression responses and thus identifies AHR among a key set of proteins responsible for mediating interactions between living organisms and foreign molecules. While extensive biochemical investigations on the interaction of AHR with ligands have been carried out, studies comparing the abilities of specific computational algorithms in explaining the potency of known AHR ligands are lacking. In this study we use molecular dynamics simulations to identify a physically realistic conformation of the AHR that is relevant to ligand binding. We then use two sets of existing data on known AHR ligands to evaluate the performance of several docking and scoring protocols in rationalizing the potencies of these ligands. The results identify an optimum set of protocols that could prove useful in future AHR ligand discovery and design as a target or anti-target. Exploration of the details of these protocols sheds light on factors operating in modeling AHR ligand binding.

The aryl hydrocarbon receptor (AHR) is one of the major xenobiotic receptors in living organisms which plays a key role in responding to foreign substances such as environmental toxins.[1,2] It plays an important role in cell differentiation and regulation. The AHR is a well-known target for the halogenated poly-aromatic hydrocarbon TCDD and is believed to be responsible for the detrimental effects of this toxin through its activation of genes for drug metabolizing enzymes like CYP1A1 and CYP1B1.[1,3,4] Diverse ligands agonize AHR and cause its translocation to the nucleus from the cytoplasm where it interacts with several co-chaperones and binds to specific response elements for metabolizing enzymes.[5–7]

While AHR biochemistry has been extensively studied, structural information about this protein is lacking. X-ray crystal structures of the AHR have not yet been obtained and structural knowledge has thus depended on homology modeling of the AHR. However, homology modeling efforts have also been few in number[8,9] and have focused on using a single methodology for docking and structure-based analysis. To our knowledge, there have been no studies exploring the ability of different scoring functions to rank binders as well as investigating alternative modes of agonist binding in the active site. In this paper we report the construction of a homology model of the AHR and perform molecular dynamics (MD) calculations on the model to pre-select an alternative physically realistic conformation for studying the molecular interaction of the AHR with known binders. Using two sets of published AHR agonists, we used two docking algorithms and five scoring functions along with a post-docking method (MM-GBSA) to validate binding data for the AHR agonists. Our results suggest the use of the Astex ASP scoring function and the MM-GBSA post-docking protocol as possible predictive tools for investigating binding of AHR agonists and induction of CYP1A1.

We used the previous template[8] of the basic helix-loop-helix protein HIF 2α which has a sequence identity of about 30% to the AHR to construct a homology model of AHR using Prime (v. 2.2). Manual sequence alignment was performed to especially align the backbone beta sheets; the aligned sequence is displayed in Figure 1b. Mutagenesis studies have identified several residues like F289, Y316, I319, F345 and A375 which have been shown to be crucial for TCDD binding.[8,10,11] In our homology model, all these residues were located in the binding pocket (Fig. 1, top).

To investigate docking and scoring procedures, we needed to locate a suitable dataset of agonists and their potencies. For this purpose the ligand binding data obtained by Hu et al.[12] in their study of CYP1A1 induction by AHR was used. From this study a set of 18 ligands was chosen[13] which had direct AHR binding assay data available. The compounds are all commercial drugs and are diverse in structure. Since we wanted to investigate the ability of the programs to at least qualitatively differentiate the compounds based on potencies, the compounds chosen were representative of ligands with high, low and medium potencies. The potency

* Corresponding author. Tel.: +1 908 231 5837.
E-mail address: Ashutosh.Jogalekar@sanofi-aventis.com (A.S. Jogalekar).

0960-894X/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2010.09.019

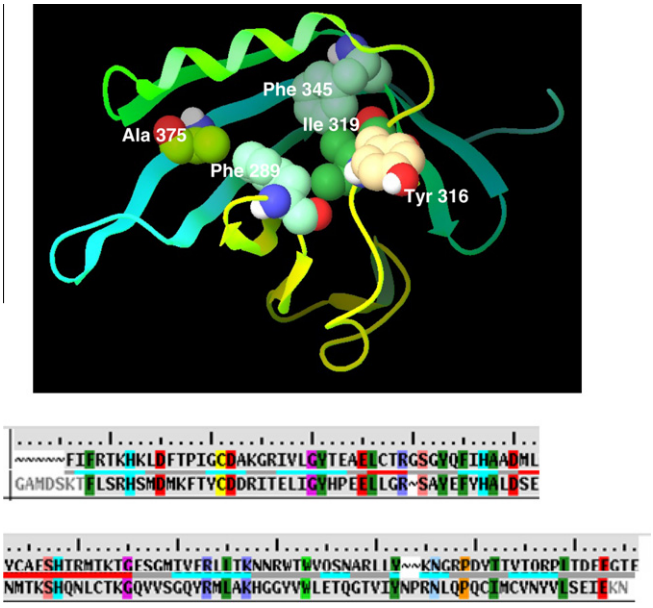

Figure 1. Top: AHR homology model with key residues. **1**, bottom: sequence alignment of AHR (top sequence) with HIF 2α template (bottom sequence).

values spanned the range of 8–111% relative to TCDD binding, which was defined as 100%. The compounds in the binding assay had also tested positive in a luciferase reporter gene assay and gel-shift assay.

Once the homology model was obtained we decided to investigate binding of the Hu et al. ligands in the active site. The prototypical standard was TCDD since its binding has been extensively investigated through mutagenesis experiments and molecular modeling.[8,10] A past study by Bisson et al. located TCDD in the binding pocket of AHR and rationalized its binding pose through postulation of proximity of residues as deduced through site directed mutagenesis data.[8] However, TCDD is a relatively small, flat, hydrophobic ligand and one could envisage other potential poses which could retain contact with the important binding site residues. To this end we decided to explore other poses of TCDD.

Glide docking of TCDD delivered a best ranked pose which by inspection mirrored the geometry of the previous pose postulated by Bisson et al.[8] In order to explore possible movement of the ligand in the binding pocket as well as relieve steric contacts from rigid docking, we decided to run 10 ns molecular dynamics simulation on this complex to facilitate movement of ligand and side chains. The MD program Desmond[14,15] with default settings was used to simulate the complex. For the simulations, the protein–ligand complex was enclosed in a cubic box filled with 5377 SPC[16] water molecules. Three Cl^- ions were added to maintain electrical neutrality. After 200-steps of minimization and equilibration, MD simulations were run under NPT conditions at 300 K for 10 ns. A 2 fs time step was used, pressure was maintained by a Martyna–Tobias Klein pressure bath[17] and temperature was maintained by a Nose–Hoover thermostat.[18] Electrostatic interactions were calculated using a smooth Particle Mesh Ewald (PME) algorithm[19] with a cutoff radius of 9 Å. The simulation analysis script in Maestro was used to analyze the results. We were interested to note that after only 1.1 ns, the ligand transitioned to a new binding mode where it remained stable for the remainder of the simulation. The movement was significant and was marked by a tilting of roughly 45° in the plane about the ligand's longest axis and about 20° about the short O–O axis, forming a distinct π-stacking interaction with a phenylalanine (Phe 289) whose side chain rotated to orient itself parallel to the ligand's aromatic ring (Fig. 2).

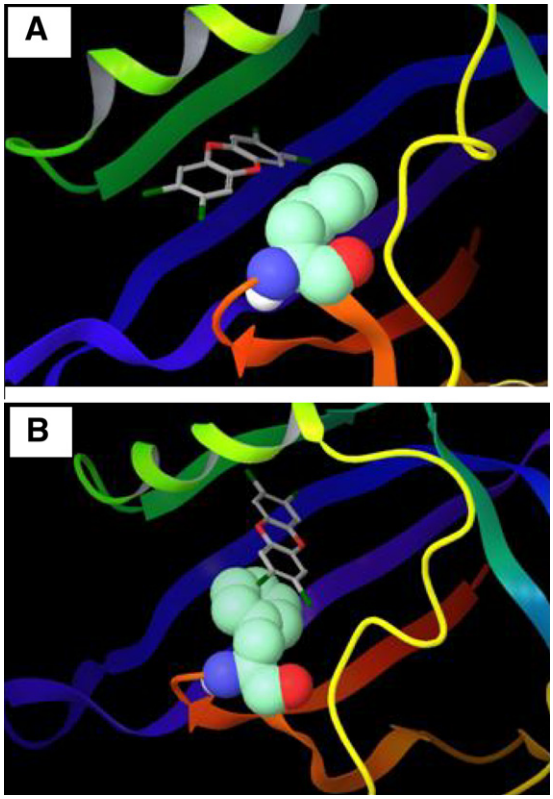

Figure 2. Snapshots from MD simulation of docked AHR–TCDD complex after 500 ps (A) and 4 ns (B). Note the movement of the ligand and Phe 289 after 4 ns to form a stacking interaction.

The physical viability of the stacked interaction is bolstered by the fact that Phe 289 has been demonstrated to be key to TCDD binding[10]; indeed, mutation of this residue to alanine completely abrogates TCDD binding. Furthermore, mutation to tyrosine reduces binding by 25-fold, which would be consistent with the weaker stacking of the relatively electron deficient aromatic ring of tyrosine compared to phenylalanine. After the ligand moved closer to the phenylalanine its position was maintained for the remainder of the simulation as shown in Figure 2. Support for the stacked pose also came from previous studies which docked TCDD into homology models of AHR and other proteins and postulated similar ligand orientations that were consistent with experimental data.[9,20–22] Furthermore, these studies have emphasized the importance of ligand planarity in influencing potent AHR binding.[21,23,24] In addition, the present arrangement duplicates contacts with amino acids mentioned earlier as important determinants of ligand binding (Fig. 1).

The observation of the viable alternative stacked pose for TCDD led us to select the protein conformation corresponding to this pose for further analysis. For this purpose, the average protein conformation between 2 and 10 ns was selected since it accurately represented the stacked ligand orientation in question; residues in the binding site showed negligible movement in these last 8 ns, increasing confidence in the choice of this conformation. Initial Glide and GOLD docking of TCDD into this protein conformation duplicated the docked poses (RMSD 0.8 Å), validating the feasibility of these programs for docking into the AHR homology model. Both programs were then used to dock the 17 other ligands from the Hu et al. study into the chosen protein structure. Glide docking was investigated in the standard (SP) and extra precision (XP) modes while for GOLD, three different scoring functions—GoldScore, ChemScore and Astex ASP score—were utilized

separately. The top scoring poses for each ligand were generated. In addition Prime MM-GBSA re-scoring was used to rescore the poses from the five docking runs. In the past this protocol has been used for productively re-scoring hits that may be ranked unfavorably by docking programs during virtual screening.[25,26]

Interestingly, the top ranked poses for the various docking programs revealed that π–π stacking between aromatic rings in the ligands and Phe 289 was a consistent interaction. Twelve out of 18 ligands displayed this contact. The stacking interactions are seen to be of the almost equienergetic parallel and parallel-displaced type, both of which have been shown to be primary π–π interaction motifs.[27,28] The AHR binding pocket is relatively hydrophobic and none of the ligands displayed hydrogen bonding interactions with the residues in the pocket. The calculated log *P* values of the ligands by themselves do not explain the potency since their correlation with the potencies was negligible. Thus, this stacking interaction can be rationalized as an important driving force for ligand binding.

Since scoring functions with their divergent energetic and parameterization criteria have been generally shown to be of limited utility in the prediction of binding energies,[29–31] our goal was to investigate the utility of various scoring functions in qualitatively ranking the ligands in terms of their potencies in the binding pocket. Based on correlation coefficients, it became apparent that the Astex ASP scoring function in combination with Prime MM-GBSA re-scoring fared best. The correlation coefficient for Astex ASP scoring increases from −0.01 to −0.56 after MM-GBSA re-scoring. In general, the MM-GBSA protocol improves the ranking for all scoring functions.

Given the error bars in the experimental dataset and the limited and diverse set of ligands used for docking, we were encouraged by these results. However, MM-GBSA has been shown to work best for congeneric series of molecules[26] of the kind more commonly investigated in lead optimization (LO) programs while the Hu et al. data set consists of structurally diverse compounds likely to be encountered in a virtual screening (VS) campaign. Thus, we were interested in applying the previous results to a more LO-like set of ligands. For this purpose, we located a study by Karplus et al. who have published a QSAR investigation utilizing three sets of 73 AHR ligands for which potency values from binding assays are available.[32] The binding assays were based on competition experiments with tritiated TCDD and the potency values expressed as pIC_{50} ranged from 3.0 to 9.3.[21] We decided to use this comprehensive set of 73 ligands to validate the protocol. The ligands chosen for the study are much more hydrophobic than the previous set and belong to three categories; dioxins (series A), dibenzofurans (series B) and biphenyls (series C), with the compounds in each set differentiated by various halogen substitution patterns on the phenyl rings. All three ligand scaffolds belong to environmental toxins which are known to activate the AHR, leading to adverse biological effects.[21]

The same protocols as before were followed for docking and scoring the Karplus et al. data set. Initial results using the scoring functions were relatively disappointing, with correlation coefficients of −0.17, −0.26 and 0.26, respectively, for SP, XP and ASP protocols. However, the utility of MM-GBSA became visible when MM-GBSA re-scoring significantly improved the results (Fig. 3). Thus, the correlation coefficients for SP, XP and ASP increased to −0.50, −0.61 and −0.56, respectively, after re-scoring.

In this case, while the XP scoring function displayed the best correlation (−0.61) after MM-GBSA re-scoring, the most significant improvement in terms of differences in scores was for the ASP score (0.26 to −0.56).

Structurally, since most ligands were planar, their poses were very similar to that for TCDD. However, the protocol also satisfactorily assigned low scores to biphenyl ligands with substituents at

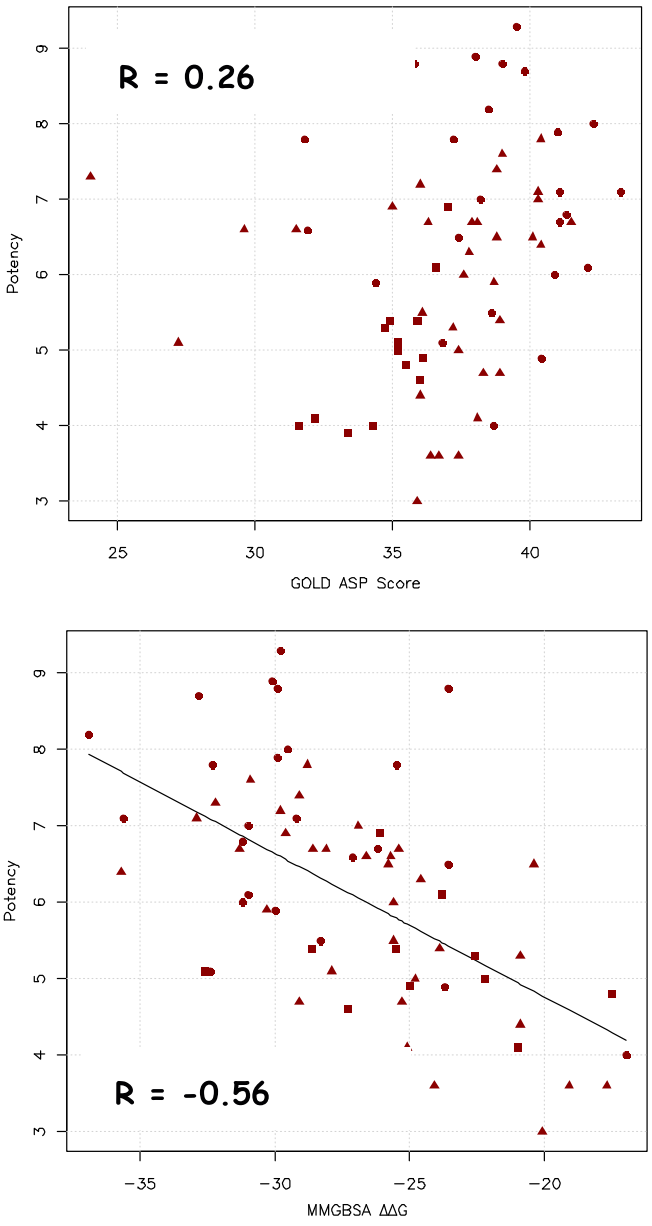

Figure 3. Plots and correlation coefficients for the Karplus et al. data set scored with Astex ASP score (top) and Astex ASP + Prime MM-GBSA (bottom). Series A: circles, series B: triangles, series C: squares.

the 2 or 2′ positions which would be expected to disrupt planarity through steric effects. A comparison of the SAR data with the MM-GBSA scores indicates that increasingly decorating the peripheral positions of the ligands with halogen atoms generally leads to an increase in potency and scores. It is observed that increasing halogen substitution embeds the flat aromatic ligands deeper into the AHR active site, facilitating improved Van der Waals contacts with certain amino acids, for instance Ala 375 and Ile 319. While this could correlate with an increase in potency, improved binding could also result from the displacement of water molecules by the added halogen atoms, a factor whose role in improving potency is now well-established.[33] The potency increase was also not caused by an increase in *C* log *P* alone, since calculated *C* log *P* values correlated very poorly with the potency values (correlation coefficient 0.04). The precise elucidation of factors contributing to potency would need more detailed investigation; however, the above results were especially encouraging since they testified to

the ability of MM-GBSA to reasonably distinguish between congeners that differed little in their structural characteristics. Thus, MM-GBSA could be potentially used as a tool in AHR ligand lead optimization.

We have built a homology model of AHR and used it to validate binding data for two different datasets of diverse ligands. A physically realistic protein–ligand configuration from a molecular dynamics simulation was selected for docking the binders. We found MM-GBSA re-scoring to be of significant utility in improving the ranking of the ligands and found the combination of the Astex ASP score and MM-GBSA re-scoring to be a particularly useful protocol for ranking such ligands. It is also worth noting that the second dataset used in the study included prominent environmental toxins, especially of the dioxin family, and the modeling investigation can thus be of potential use in exploring SAR relationships in this set of environmentally important compounds. To our knowledge, studies validating docking programs on the AHR have not been carried out, and it is hoped that the above combination of docking and scoring would be a possible predictive tool for investigating AHR binding and CYP1A1 induction.

References and notes

1. Denison, M. S.; Nagy, S. R. *Annu. Rev. Pharmacol. Toxicol.* **2003**, *43*, 309.
2. Denison, M. S.; Pandini, A.; Nagy, S. R.; Baldwin, E. P.; Bonati, L. *Chem. Biol. Interact.* **2002**, *141*, 3.
3. Okey, A. B.; Riddick, D. S.; Harper, P. A. *Toxicol. Lett.* **1994**, *70*, 1.
4. Riddick, D. S.; Huang, Y.; Harper, P. A.; Okey, A. B. *J. Biol. Chem.* **1994**, *269*, 12118.
5. Petrulis, J. R.; Perdew, G. H. *Chem. Biol. Interact.* **2002**, *141*, 25.
6. Ikuta, T.; Eguchi, H.; Tachibana, T.; Yoneda, Y.; Kawajiri, K. *J. Biol. Chem.* **1998**, *273*, 2895.
7. Pollenz, R. S.; Barbour, E. R. *Mol. Cell Biol.* **2000**, *20*, 6095.
8. Bisson, W. H.; Koch, D. C.; O'Donnell, E. F.; Khalil, S. M.; Kerkvliet, N. I.; Tanguay, R. L.; Abagyan, R.; Kolluri, S. K. *J. Med. Chem.* **2009**, *52*, 5635.
9. Jacobs, M. N.; Dickins, M.; Lewis, D. F. *J. Steroid Biochem. Mol. Biol.* **2003**, *84*, 117.
10. Pandini, A.; Soshilov, A. A.; Song, Y.; Zhao, J.; Bonati, L.; Denison, M. S. *Biochemistry* **2009**, *48*, 5972.
11. Goryo, K.; Suzuki, A.; Del Carpio, C. A.; Siizaki, K.; Kuriyama, E.; Mikami, Y.; Kinoshita, K.; Yasumoto, K.; Rannug, A.; Miyamoto, A.; Fujii-Kuriyama, Y.; Sogawa, K. *Biochem. Biophys. Res. Commun.* **2007**, *354*, 396.
12. Hu, W.; Sorrentino, C.; Denison, M. S.; Kolaja, K.; Fielden, M. R. *Mol. Pharmacol.* **2007**, *71*, 1475.
13. The 18 ligands chosen from Ref. 12 were methylcholanthrene, n., naphthoflavone a, TCDD, leflunomide, pioglitazone, flutamide, valsartan, nimodipine, indomethacin, aconitine, anastrozole, benzethonium, bezafibrate, fluvastatin, omeprazole, pyrilamine and sertraline.
14. Bowers, K. J.; Dror, R. O.; Shaw, D. E. *J. Chem. Phys.* **2006**, *124*, 184109.
15. Shaw, D. E. *J. Comput. Chem.* **2005**, *26*, 1318.
16. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, 91.
17. Martyna, G. J. *Phys. Rev. E Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1994**, *50*, 3234.
18. Evans, D. J.; Holian, B. L. *J. Chem. Phys.* **1985**, *83*, 4069.
19. Darden, T.; Perera, L.; Li, L.; Pedersen, L. *Structure* **1999**, *7*, R55.
20. McKinney, J. D.; Chae, K.; Oatley, S. J.; Blake, C. C. *J. Med. Chem.* **1985**, *28*, 375.
21. Safe, S. H. *Annu. Rev. Pharmacol. Toxicol.* **1986**, *26*, 371.
22. Lewis, D. F.; Jacobs, M. N.; Dickins, M.; Lake, B. G. *Toxicology* **2002**, *176*, 51.
23. Fujita, Y.; Yonehara, M.; Tetsuhashi, M.; Noguchi-Yachide, T.; Hashimoto, Y.; Ishikawa, M. *Bioorg. Med. Chem.* **2009**.
24. Copeland, K. L.; Anderson, J. A.; Farley, A. R.; Cox, J. R.; Tschumper, G. S. *J. Phys. Chem. B* **2008**, *112*, 14291.
25. Graves, A. P.; Shivakumar, D. M.; Boyce, S. E.; Jacobson, M. P.; Case, D. A.; Shoichet, B. K. *J. Mol. Biol.* **2008**, *377*, 914.
26. Guimaraes, C. R.; Cardozo, M. *J. Chem. Inf. Model.* **2008**, *48*, 958.
27. Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 10887.
28. Aravinda, S.; Shamala, N.; Das, C.; Sriranjini, A.; Karle, I. L.; Balaram, P. *J. Am. Chem. Soc.* **2003**, *125*, 5308.
29. Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. *J. Med. Chem.* **2006**, *49*, 5912.
30. Perola, E.; Walters, W. P.; Charifson, P. S. *Proteins* **2004**, *56*, 235.
31. Wang, R.; Lu, Y.; Fang, X.; Wang, S. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2114.
32. So, S. S.; Karplus, M. *J. Med. Chem.* **1997**, *40*, 4360.
33. Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. *J. Am. Chem. Soc.* **2008**, *130*, 2817.