J Comput Aided Mol Des (2013) 27:707–721
DOI 10.1007/s10822-013-9670-6

Experimental versus predicted affinities for ligand binding to estrogen receptor: iterative selection and rescoring of docked poses systematically improves the correlation

James S. Wright · James M. Anderson · Hooman Shadnia · Tony Durst · John A. Katzenellenbogen

Received: 2 May 2013 / Accepted: 2 August 2013 / Published online: 24 August 2013

Abstract The computational determination of binding modes for a ligand into a protein receptor is much more successful than the prediction of relative binding affinities (RBAs) for a set of ligands. Here we consider the binding of a set of 26 synthetic A-CD ligands into the estrogen receptor ERα. We show that the MOE default scoring function (London dG) used to rank the docked poses leads to a negligible correlation with experimental RBAs. However, switching to an energy-based scoring function, using a multiple linear regression to fit experimental RBAs, selecting top-ranked poses and then iteratively repeating this process leads to exponential convergence in 4–7 iterations and a very strong correlation. The method is robust, as shown by various validation tests. This approach may be of general use in improving the quality of predicted binding affinities.

Keywords Docking · Scoring · Iterative rescoring · Estrogen receptor

Introduction

It is well known that docking a set of ligands into a protein receptor can usually identify the important binding modes, but prediction of relative binding affinities (RBAs) has been much less successful [1–5]. As an example, a very thorough study by Warren et al. in 2006 [1] using various methodologies, and with a large diverse set of ligands and receptors, found only very weak correlations between measured and calculated binding affinities. In the present paper we explore one methodology which may lead to an improved prediction of binding affinities. Here our focus will be on the evaluation of scoring functions used to rank-order the docked poses (DPs), and examine their optimization.

A great deal of development is being done in several areas which may indirectly lead to improved scoring functions (SFs). These include the treatment of receptor flexibility [6], the use of multiple receptor geometries [7, 8], how to recognize optimum ligand alignment [9, 10], estimation or calculation of entropic effects[11], estimating solvation energies and hence ligand desolvation penalties [12], whether or not to include bridging water molecules [6, 13], and how to best determine protonation states [14] of residues. The growing popularity of consensus scoring [15, 16], in which a variety of different scoring functions are combined and averaged, is clearly an attempt to bring some kind of representative average out of an otherwise confusing situation.

With respect to virtual screening of large databases, in the present work we will assume that a rapid screening method has been applied to determine "hits" and that a

Electronic supplementary material The online version of this article (doi:10.1007/s10822-013-9670-6) contains supplementary material, which is available to authorized users.

J. S. Wright (✉) · J. M. Anderson
Department of Chemistry, Carleton University, 1125 Colonel By Dr., Ottawa K1S 5B6, Canada
e-mail: jim_wright@carleton.ca

H. Shadnia
Chemical Computing Group, 1010 Sherbrooke St. W., Montreal, QC H3A 2R7, Canada

T. Durst
Department of Chemistry, University of Ottawa, D'Iorio Hall, 10 Marie Curie St., Ottawa K1N 6N5, Canada

J. A. Katzenellenbogen
Department of Chemistry, University of Illinois, Urbana, IL 61801, USA

small subset of ligands has been selected for lead optimization. This subset, which in general may only consist of 10–100 ligands, will then become the focus for determination of relative binding affinities. In this paper we report on a scoring procedure which we have found to be very promising when applied to ligand binding to the protein estrogen receptors subtype alpha (ERα) and beta (ERβ). The procedure is based on a force-field methodology which incorporates a multiple linear regression (MLR) analysis, leading to the iterative improvement of a scoring function by reselection of preferred poses from the same (large) set of docked poses (DPs). Once the DPs have been established, which is CPU-time intensive, sorting and iterating through the docked poses to optimize the scoring function is very fast.

Various types of iterative pose selection have been described previously in the literature. Early work by Warshel, Aqvist and coworkers [17–20] explored the concept of using feedback from binding affinity measurements to improve the shape of the potential function for the force field. Huang and Zhou [21] also used an iterative approach to derive interaction potentials for ligand–protein interactions. Neural networks incorporating feedback loops have been extensively used by Jain and coworkers [22–24] and others [25, 26] and form part of several software packages. Machine learning is another variant which attempts through various learning modes to improve predictive power of scoring functions [27–29].

An iterative approach to rescoring docked poses, similar in concept to the present work, was reported by Jain in 1996 [23]. In that paper he refers to earlier work [24] on the Compass algorithm with the following description: "The training algorithm iterates parameter estimation and ligand pose optimization. Convergence to a stable set of parameters occurs within five iterations". However, the work closest in spirit to the present application was developed by Martin and Sullivan [30, 31] who discussed all the basic methodology of an iterative approach to improving scoring functions using their Autoshim software. In that work they discuss of use of "shims" which can be considered as modifications to their potential energy function. These authors used feedback from measured IC_{50} values for a series of kinase inhibitors to adjust the magnitude of the shims, and in doing so significantly improved the correlation with experimental data, as measured by the cross-validated correlation coefficient q^2 [30]. Their iterative approach maximizes the scoring function and thereby selects the best docked pose from the hundreds of DPs generated by docking for each ligand. They also test and discard an iterative approach which minimizes the deviation from experimental IC_{50} values. Finally, their iterative method converged in 2–4 iterations in general and were independent of the choice of starting "best" DPs.

In our opinion the suitable choice of a scoring function is one of the most important problems under scrutiny in computational drug design, and the use of iterative methods to improve scoring functions represents a potentially important and under-utilized path towards this goal. However, one problem sure to be encountered by the potentially interested user is that the methods described in the literature are intimidating in their complexity. For example, the Autoshim method [30] introduces >100 descriptors involving the shim location and magnitude along with many other descriptors, and as a result it is hard to interpret the significance of any single descriptor. In the present paper we describe in detail our methodology based on MOE [32], a widely used software platform, and use only four descriptors to obtain correlation coefficients $r^2 > 0.9$ for a data set containing 17 ligands. A very detailed set of on-line information has been provided so that the entire methodology could be reproduced by a motivated reader. As will be seen from our own results, the procedure for iterative pose selection is not only simple and rapid, once a set of docked poses is obtained, but it is also robust and not prone to overfitting, in agreement with the results of Martin and Sullivan [30], even though our implementation of the descriptor set is quite different from theirs.

Methodology

A-CD ligands

In two recent papers [33, 34] we described the synthesis and binding affinity to the estrogen receptor ERα for a novel set of flexible A-CD ligands. The compounds were synthesized by combining a functionalized phenolic A-ring through a coupling reaction with a previously prepared CD fragment. As shown in Fig. 1 below, the stereochemistry at the C-D ring junction, i.e. between C13 and C14 using notation derived from the standard steroid ring numbering, is different in A-CD from that in estradiol (E2) [33, 34], but is the same at all other stereo centers. Here we use the term "*cis*" when the H-atom at C14 is on the same side of the CD plane as the methyl group at C13, and "*trans*" when it is on the opposite side as in E2 (see Fig. 1). Substituents, which include F, Cl, CH_3 and CF_3, are located at positions 2, 4 and 5 on the A-ring. Compounds 1a-1 l and 1o have a hydroxyl group at C17, whereas compound 1p has a carbonyl group (similar to estrone). Compound 1q, an analog of the popular hormone supplement ethynyl estradiol, contains both the 17-OH group and the17-ethynyl group. The set of ligands also includes three compounds with inverted chirality at C9 (2a, 2e, 2f), four compounds with a double

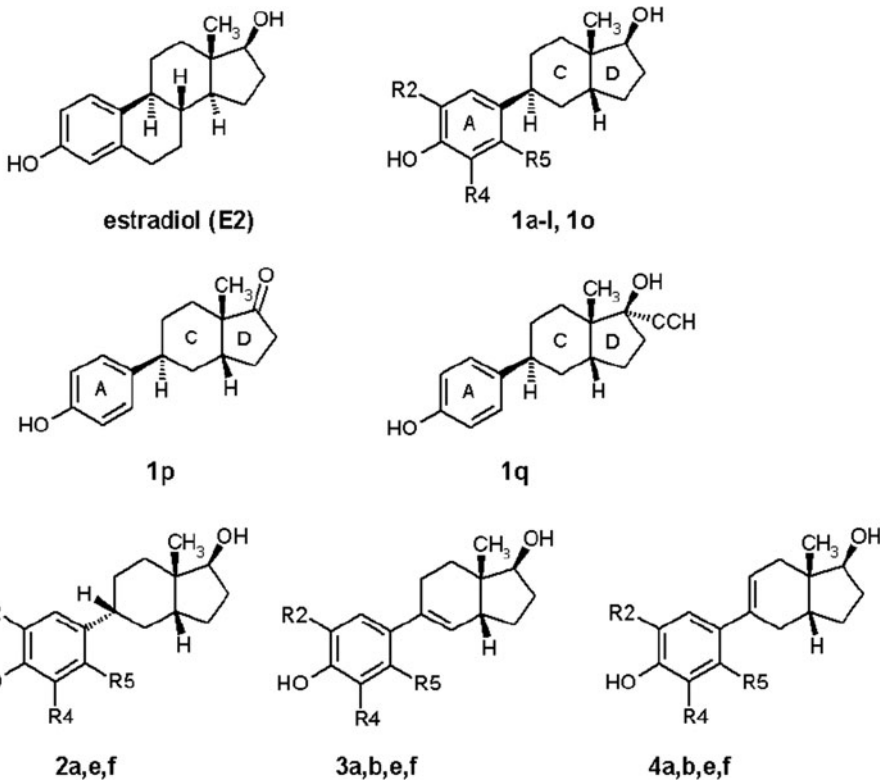

Fig. 1 Numbering scheme for A-CD compounds. Substituents are in the order R2, R4, R5, where R2 is the substituent located at position 2 in A-ring, R4 = substituent at position 4, etc. If R5 = H, due to free rotation about the A-CD bond, positions R2 and R4 are equivalent. Phenolic OH is at position 3

bond in the C-ring away from the methyl group (3a, 3b, 3e, 3f) and four with a double bond in the C-ring toward the methyl group (4a, 4b, 4e, 4f). Except for compounds 4a and 4b, synthetic details and RBA values for all compounds in the 2-, 3- and 4-series have not been published previously; the missing synthetic details will be reported in a separate publication along with an expanded set of A-CD ligands.

Relative binding affinities

All of the experimental RBAs were obtained using the same competitive radioactive tracer procedure using tritiated estradiol [33, 34] and therefore form a self-consistent set. They are listed in Table 1 along with the substituents on the A-ring at positions 2, 4 and 5. Inspection of Table 1 shows that the RBAs range over almost 5 orders of magnitude and hence provide a sufficient range of affinities to provide a good test for any scoring function. As discussed in previous work [34] location of substituents at the 2 and 4 positions decrease the RBA whereas substituents at the 5-position increase it; e.g. compound 1f has an RBA approaching that of the reference compound E2 (89.7 vs. 100 %). Compounds containing inverted chirality at C9, i.e. 2a, 2e, 2f, are quite inactive, whereas several compounds (3b, 3f, 4f) with an unsaturated C-ring have an RBA exceeding that of E2.

Prediction of the RBA

In previous work [33, 34] which covered some of the ligands in the Training Set we used a force field-based scoring function constructed from the interaction energy E_{int} in the bound complex to rank-order the docked poses (DPs). The scoring function was therefore similar to the one used here in the following discussion, but based on a different starting point (E_{int}) and did not use iterative pose selection. This approach was successful for the set of A-CD ligands, but was somewhat lacking in generality. We are currently in the process of using a similar scoring methodology to that described here and applying it to sets of *cis*- and *trans*- ACD ligands (unpublished data). Our present approach is based on use of force field-based terms in an attempt to compute the free energy change for dissociation of ERα containing a bound ligand, i.e. for the reaction Complex → Ligand + Receptor. The reaction has $K_d = 0.2$ nM [34] and the free energy change is therefore positive by 13.2 kcal/mol at 25 C. Essential features of this approach, which have been discussed recently by Merz [35] and others involves calculation for each ligand of the following force-field terms: (1) minimum-energy conformer of the free ligand, E_L; (2) energy of the distorted ligand in the complex, E_L'; (3) energy of the prepared receptor in solvent, E_R; (4) energy of the distorted receptor in the complex, E_R'; (5) interaction energy between ligand

Table 1 Experimental values of RBAs) for the full set of A-CD ligands bound to ERα [see also Ref. [34]. Reference value: (E2, 100 %); error bars in parentheses

Compound	R2	R4	R5	RBA (%)
E2	–	–	–	100
1a	H	H	H	1.47 (0.26)
1b	H	F	H	1.04 (0.20)
1c	H	H	F	27.3 (0.69)
1d	H	H	Cl	52.3 (12.0)
1e	H	H	Me	2.82 (0.45)
1f	H	H	CF3	89.7 (14.0)
1 g	F	F	H	0.040 (0.002)
1 h	Cl	Cl	H	0.004 (0.001)
1i	H	F	F	4.62 (0.93)
1j	F	H	F	0.380 (0.090)
1 k	F	F	F	0.186 (0.010)
1 l	H	Me	F	1.75 (0.50)
1o	H	Me	Me	0.065 (0.001)
1p	H	H	H	0.008 (0.001)
1q	H	H	H	0.088 (0.006)
2a	H	H	H	0.061 (0.002)
2e	F	F	H	0.006 (0.001)
2f	Cl	Cl	H	0.004 (0.001)
3a	H	H	H	0.246 (0.02)
3b	H	H	Cl	195 (35)
3e	F	H	F	0.038 (0.005)
3f	H	H	CF3	189 (1.0)
4a	H	H	H	0.415 (0.08)
4b	H	H	Cl	59.9 (7.9)
4e	F	H	F	0.436 (0.08)
4f	H	H	CF3	122 (17)

and receptor in complex, E_{int}. The change in energy of the ligand on binding is therefore $\Delta E_L = E'_L – E_L$ and similarly for the receptor $\Delta E_R = E'_R – E_R$ (for further details see Computational Methods section).

There is one additional term which is the (de)solvation penalty, i. e. (6) the energy required to take the ligand out of solution and put it into the complex, ΔG_{solv}. It is estimated as the free energy of the ligand in aqueous solution and is calculated with Gaussian'03 software [36] using the conductor-like polarisable continuum model (CPCM) on DFT-optimized global minimum energy geometries of the ligand (Keywords: [HF/6-31G(d), SCRF = (CPCM, Read, Solvent = Water)].

The resulting expression for the energy change in the reaction Complex → Ligand + Receptor thus can be written $\Delta E = C_0 + C_1(E_{int–}) + C_2(\Delta E_L) + C_3(\Delta E_R) + C_4(\Delta G_{solv})$. The procedure is therefore the following for any given ligand: (1) using MOE2012 [32] or updated versions, generate a database of ligand conformers which lie below 7 kcal/mol, (2) create a docking box at least 1 Å larger in every direction than the ligand, (3) generate a set of starting poses by choosing random translation, rotation, and orientation of rotatable bonds within the docking box to create ca. 500–5,000 poses, (4) dock the poses into the prepared receptor allowing the receptor to relax in the process, and saving a set of DPs, and (5) for each DP evaluate the terms $E_{int–}$, ΔE_L and ΔE_R. ΔG_{solv} is evaluated only once, for the lowest-lying conformer of each ligand based on E_{int}. Next, (6) we must choose a scoring function so that the highest scoring docked pose for each ligand can be determined. This will lead to a new database containing ligand entries, one per ligand, containing the ligand name, the measured RBA value, and values for the four energy descriptors. Finally, the last step is to do an MLR to determine the coefficients C_0–C_4 by partial least-squares (PLS) fitting to the experimental RBAs.

A crucial step in the above procedure is the choice of scoring function to rank-order the DPs. However, before making a choice of any particular function, we consider some general characteristics of ligand binding, using known features of E2 in ERα as an example.

Protein–Ligand H-bonding Network

For the naturally occurring co-crystallized ligand E2 in ERα, the crystal structure (1GWR) suggests what should be found for an optimal binding mode between ligand and receptor. Of course H-atoms are not visible in the crystal structure but after addition of H-atoms it is clear what is the preferred mode of binding. The phenolic OH group at C3 participates in a hydrogen bonding network which includes the bridging water molecule HOH2009 and the residues Glu353 and Arg394. Five distinct hydrogen bonds are formed, as shown in Fig. 2a. At C17, the hydroxyl group forms a single hydrogen bond to His524, shown in Fig. 2b. Other variations of this H-bonding network are possible, and Protonate3D [14] is a program in MOE which considers the local environment before deciding on the proper protonation state. However, in this paper we will use the simpler placement of H-atoms in default geometries, followed by energy minimization.

Energy-based terms for dissociation of bound complex

As stated above, there are several energy-based properties useful for defining an optimal structure. The first such indicator is the interaction energy, E_{int}. A more negative E_{int} indicates a stronger binding interaction between ligand and receptor. For example, in the case of the co-crystallized ligand E2, E_{int} is ca. −74 kcal/mol (MMFF94 s force field [37]). Since E2 is known to be a very strong binder with subnanomolar binding affiinity, it is expected that A-CD ligands with comparably strong binding would also show similar values for E_{int}. The next such indicator is the

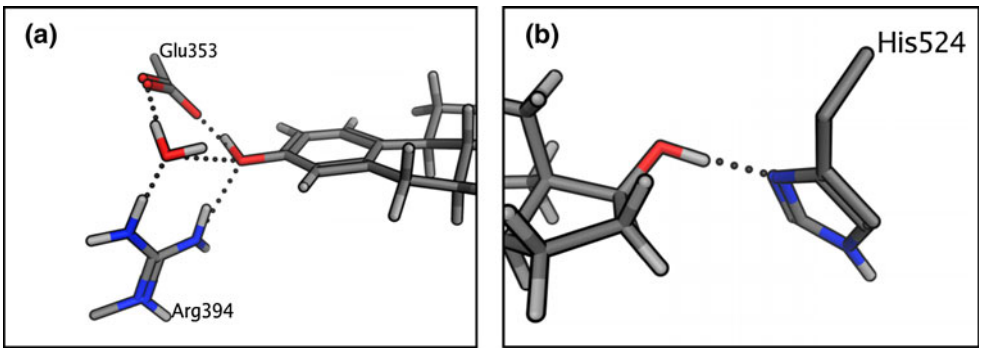

Fig. 2 Hydrogen-bonding network for 17β-estradiol bound to ERα, starting from the crystal structure (pdb code: 1GWR) and manually optimizing the orientation of the bridging water and the ligand OH bonds. **a** Network surrounding the 3-OH group on the A-ring. **b** Network surrounding the 17-OH group on the D-ring. For clarity, only residues having polar interactions with the active site are shown

deformation energy of the ligand in the bound complex, ΔE_L. Here a high value indicates a strongly deformed ligand, so lower values of ΔE_L (e.g. below 10 kcal/mol) correspond to more physically reasonable (less distorted) poses. Similarly, the deformation energy of the receptor, ΔE_R, should be in the lower end of its range (less than ca. 50 kcal/mol above a reference value of 300 kcal/mol for the receptor built from 1GWR), because this corresponds to a less strained receptor. The fourth energy-related variable of interest is the solvation energy of the free ligand. Using the CPCM method for E2 gives a value of −12.40 kcal/mol. More negative values imply greater stabilization of the ligand in water and hence a greater desolvation penalty to place the ligand in the receptor, so more positive (i.e. less negative) values are favored.

Scoring function: London dG

To rank-order the docked poses, MOE at one point used as default the London dG scoring function [32]. This scoring function contains a term for the gain or loss of translational and rotational entropy on formation of the bound complex, an energy term related to freezing of ligand rotatable bonds in the bound complex, a hydrogen-bonding term for ligand–protein interactions, a metal ligation term and a desolvation penalty term for removing the ligand from aqueous solution. Coefficients for these terms were derived from fitting to ca. 400 X-ray crystal structures of protein–ligand complexes using experimental binding affinity data [see MOE documentation 2012]. The more negative the value of the London dG score, the more strongly bound is the Complex [32].

Training and test sets

To study the behavior of predicted RBA values following docking and scoring, we require a training set of ligands which form a representative set from the 27 data values listed in Table 1. In order to be representative, the training set should include compounds of high RBA (>10 %), medium RBA (0.1–10 %) and low RBA (<0.1 %). In addition, it should include representative compounds from each distinct structural type. These types can be classified as A-CD where C is saturated (1a–1q), those with inverted stereochemistry at C9 (the 2-series), and those with unsaturation in the C-ring (the 3- and 4-series). The training set should also include the reference ABCD compound E2. Parameter values obtained from the training set will then be used unchanged to predict properties of the test set, which should also be as inclusive as possible of both RBA and compound types. Beyond that, values should be chosen randomly from the full data set to create the subsets.

Signs of the coefficients

Since the Complex is bound with respect to its dissociation products Ligand + Receptor, the free energy of products is above (more positive) than that of reactants. Therefore anything which increases the free energy change of the dissociation reaction will increase the binding affinity. Consider the terms in the equation $\Delta E = C_0 + C_1(E_{int-}) + C_2(\Delta E_L) + C_3(\Delta E_R) + C_4(\Delta G_{solv})$. The interaction energy E_{int} is negative and stabilizes the complex (i.e. more negative means stronger binding), so the coefficient C_1 should also be negative. The ligand distortion energy ΔE_L is positive and destabilizing, therefore C_2 should also be negative (more ligand distortion energy, less binding). The receptor distortion energy ΔE_R is positive and destabilizing so C_3 should be negative (more receptor distortion energy, less binding). Finally, the ligand solvation energy is negative, and the more negative, the bigger the desolvation penalty, therefore to reflect this, C_4 should be positive. Any inconsistency in signs relative to the above may mean that the particular energy descriptor is not meaningful. However, some distortion of the ligand is always present in the receptor and vice

versa, so a simple linear term in ΔE_L and ΔE_R may not be an adequate representation.

Results and discussion

Training Set

Training Set 1 was chosen as described above. It contains 17 ligands and includes both the reference structure E2 and the important prototype compound 1a (for Test Set 1 which contains the other 10 ligands, see Table 6). The conformer generation and docking procedure led to 16,172 DPs for the training set and 9,186 DPs for the test set. The London dG scoring function was evaluated for all docked poses for each ligand in Training Set 1 (i.e. 16,172 evaluations). The most negative value of London dG (highest ranked) from the set of docked poses for each ligand was selected, creating a new output file, shown in Table 2. This Table shows the 17 experimental values of log RBA (expt.) and the computed London dG score, where the data entries for each ligand correspond to the top-ranked docked pose.

Once the ranking of the DPs has been established, we can compute the four corresponding energy terms E_{int}, ΔE_L, ΔE_R and ΔG_{solv}, which are also shown in Table 2. A first warning sign is that the values of E_{int} are well above the expected region near −70 kcal/mol. Some of the ligand distortion energies ΔE_L are quite high (e.g. 13.59 kcal/mol for 1c) and similarly for the receptor distortion energy ΔE_R, with one value (1o) as high as 73.8 kcal/mol. The solvation energies were assumed to be the same for all docked poses of a given ligand, so they play no role in pose selection. This is clearly a crude assumption, since internal versus external hydrogen bonds may change the solvation considerably. However, each determination of solvation energy required about 1 h CPU time on a large SUN cluster, so calculation of ΔG_{solv} for each of the 15,000 docked poses would be prohibitively expensive.

Next, coefficients of the four energy terms were fitted by MLR using the partial least squares method to give the equation for Scoring Function 1 (SF1) as

$SF1 = \text{Predicted log RBA} = 0.579 – 0.0225(E_{int}) – 0.135(\Delta E_L) – 0.0387(\Delta E_R) – 0.130(\Delta G_{solv})$, with correlation coefficient $r^2 = 0.245$, rms. dev. = 1.32 log unit and cross-validated correlation coefficient (Leave One Out method) $q^2 = 0.026$. The graph of this data fit is shown in Fig. 3a. The correlation is very weak, as expected from the very poor correlation coefficient, large standard deviation, and near-zero cross-validated correlation coefficient. Thus, it appears doubtful that SF1 can even discriminate between actives and inactives.

Figure 3b gives some insight into the problem with the DPs generated by the London dG function (and associated SF1). It can be seen that the entire ligand has been rotated by 180 deg within the active site, relative to the crystal structure 1GWR, resulting in a poor H-bond network. Other ligands in Training Set 1 have different, but also incorrect binding modes. Thus, SF1 (which scores the same docked poses given by London dG) gives a very poor

Table 2 Calculated energy descriptor set (E_{int}, ΔE_L, ΔE_R and ΔG_{solv}) and London dG Scoring Function for the 17 ligands in the Training Set 1

Ligand	log RBA (expt.)	E_{int}	ΔE_L	ΔE_R	ΔG_{solv}	London dG	SF1
E2	2.00	−53.89	6.11	61.01	−12.40	−14.39	0.22
1a	0.17	−54.32	6.29	36.54	−10.69	−13.57	0.93
1b	0.02	−52.85	6.87	37.94	−8.09	−13.93	0.43
1c	1.44	−62.04	13.59	51.44	−10.07	−14.24	−0.54
1e	0.45	−58.09	5.77	49.79	−9.58	−13.59	0.43
1f	1.95	−65.22	4.12	50.40	−8.20	−14.94	0.61
1 h	−2.40	−56.73	9.97	50.02	−5.97	−14.03	−0.65
1 k	−0.73	−59.76	7.51	56.24	−7.43	−14.63	−0.30
1 l	0.24	−56.30	7.25	49.12	−8.96	−14.42	0.14
1o	−1.19	−47.03	9.70	73.82	−8.59	−14.41	−1.41
1p	−2.10	−52.62	12.59	54.84	−11.15	−12.86	−0.60
2a	−1.21	−62.19	9.77	50.80	−11.75	−14.54	0.23
2e	−2.22	−56.94	10.84	76.86	−8.70	−15.26	−1.44
3e	−1.42	−60.49	6.50	44.74	−12.32	−14.97	0.94
3f	2.28	−57.21	8.73	40.93	−8.98	−14.69	0.28
4b	1.78	−58.88	9.17	33.51	−9.45	−14.06	0.60
4e	−0.36	−45.27	12.82	66.70	−11.67	−14.99	−1.19

Docked poses were taken from the top-ranked (most negative) values based on the London dG Scoring Function. All energy values in kcal/mol

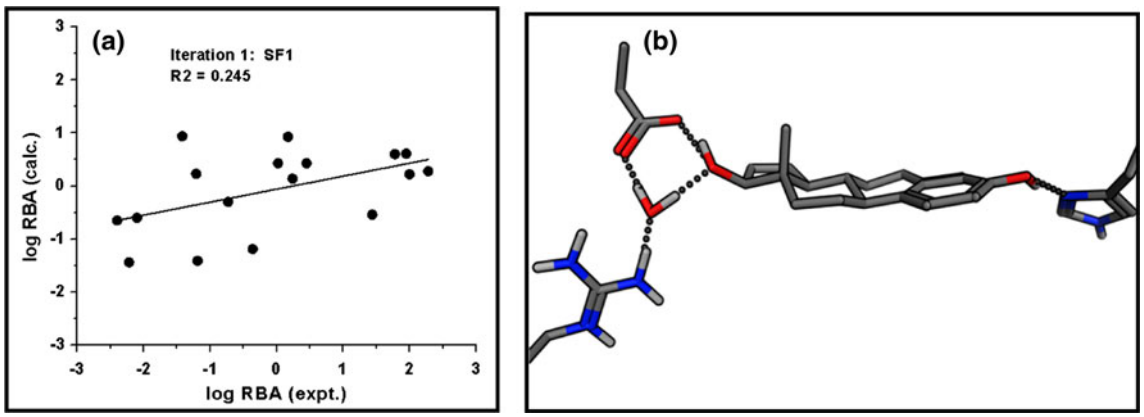

Fig. 3 **a** Initial correlation between predicted and experimental log RBA values for Training Set 1, based on a multiple linear regression over four energy terms and pose selection based on the London dG scoring function. **b** Binding mode of E2, showing the top-ranked docked pose according to the London dG scoring function

correlation with experimental RBAs, not surprising when it incorrectly predicts the best known binding mode for E2.

Nevertheless, suppose that we use SF1 as the start of a new procedure in which we seek to improve the quality of the correlation. To accomplish this, values of SF1 were calculated for all 16,172 DPs, instead of just the 17 entries in Table 2, to obtain a new rank-ordering of the poses, where the highest-ranked DPs were those with the largest value of SF1 (a more positive value of SF1 implies a larger log RBA which implies stronger binding). The top-ranked poses from this new rank-ordered set based on SF1 are different from those obtained using London dG. The new top-ranked DP for each ligand was collected and shown in Table 3, along with the corresponding newly calculated values of the four energy descriptors.

It is immediately evident from Table 3 that there has been a significant improvement in the (newly) chosen set of DPs. For example, values of E_{int} are now in the range −51 to −75 kcal/mol, compared to −47 to −65 kcal/mol (Table 2), a very substantial change for the better. Similarly, ligand distortion energies are better (< + 8 vs. < + 14 kcal/mol, Table 2) and so are receptor distortion energies (all of which are < + 40 kcal/mol vs. only seven < + 50 kcal/mol, Table 2).

Coefficients from the MLR fit to the new data set in Table 3 gave a revised SF, SF2 = −10.85 −0.223(E_{int}) −0.410(ΔE_L) −0.0446(ΔE_R) +0.0462(ΔG_{solv}), where $r^2 = 0.853$, rms. dev. = 0.582 log unit and $q^2 = 0.440$. The correlation obtained after fitting the re-ordered DPs from SF1, which is shown in Fig. 4a, has much better predictive value by all three measures, although the value of q^2 is still quite low Fig. 4b shows that the top-ranked DP for E2 now has the correct binding orientation, although the H-bonding network is not quite optimal (Arg394 is slightly twisted, losing one H-bond).

Table 3 Iteration No. 2: Docked poses for Training Set 1 were re-ranked using the SF1 scoring function

Ligand	log RBA (expt.)	E_{int}	ΔE_L	ΔE_R	ΔG_{solv}	SF2
E2	2.00	−68.11	2.95	26.36	−12.40	1.39
1a	0.17	−64.64	4.88	19.22	−10.69	0.22
1b	0.02	−60.47	5.30	27.23	−8.09	−1.12
1c	1.44	−65.51	5.00	19.59	−10.07	0.37
1e	0.45	−66.28	4.40	23.17	−9.58	0.66
1f	1.95	−71.35	6.16	22.09	−8.20	1.18
1 h	−2.40	−55.92	3.69	40.00	−5.97	−1.95
1 k	−0.73	−65.04	5.04	28.74	−7.43	−0.03
1 l	0.24	−67.41	5.13	19.48	−8.96	0.80
1o	−1.19	−64.29	6.82	33.61	−8.59	−1.19
1p	−2.10	−51.41	2.92	26.35	−11.15	−2.27
2a	−1.21	−62.64	6.64	25.69	−11.75	−1.29
2e	−2.22	−61.17	7.38	35.73	−8.70	−2.22
3e	−1.42	−62.19	4.74	20.75	−12.32	−0.41
3f	2.28	−75.31	5.61	17.26	−8.98	2.47
4b	1.78	−70.79	4.21	15.54	−9.45	2.09
4e	−0.36	−68.59	7.25	21.53	−11.67	−0.01

The most positive value of SF1 for each ligand was used to generate this table. Energy values in kcal/mol

The process of re-ranking the DPs and rescoring according to SF2 led to the next iteration SF3, after which a new MLR was performed, and the process was iterated until no further change occurred in the set of 17 top-ranked DPs. Figure 5 shows the dependence of the correlation coefficient r^2 on iteration number. It can be seen that the graph shows excellent convergence properties, with SF3 already almost converged and SF8 fully converged.

The data in Fig. 5 were fitted by an (inverted) exponential function, where the plateau occurs at the

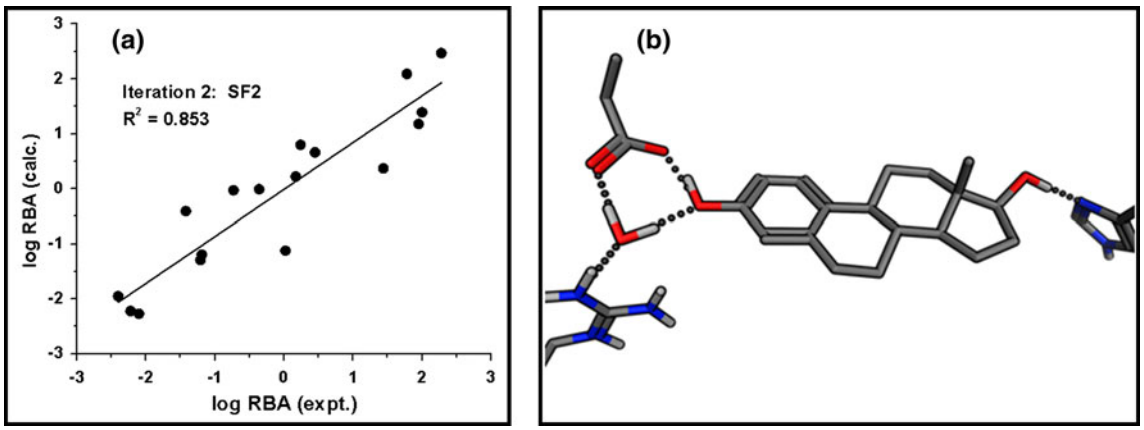

Fig. 4 **a** Correlation between the predicted log RBA obtained from the MLR, second iteration, and the experimental log RBA values for Training Set 1. **b** Top-ranked binding mode of E2, as given by the top-ranked pose from SF2

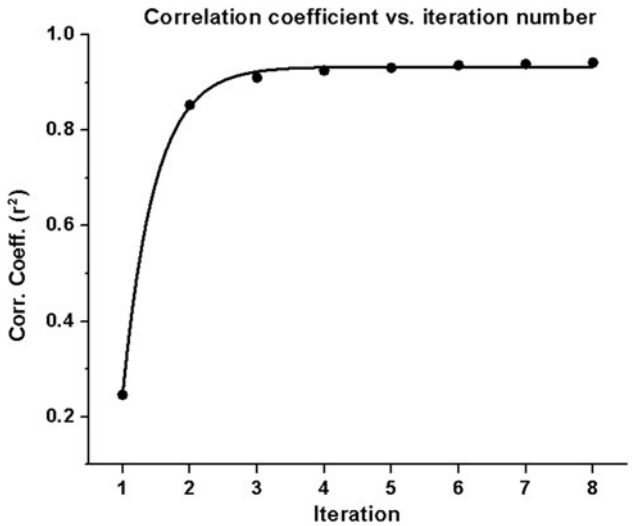

Fig. 5 Change in correlation coefficient r^2 as a function of iteration number

convergence limit near $r^2 = 0.94$. The equation of the fit is given by

$r^2 = 0.940–5.65 \exp [-2.11 \text{ (Iteration no.)}]$,

Which, as Fig. 5 shows, is a remarkably good fit to the iteration data. A similar result was obtained for the cross-validated correlation coefficient (data not shown). We shall refer to this type of convergence behavior in the correlation coefficients as "exponentially convergent", clearly a desirable property in this methodology. Table 4 shows the final (converged) values from Training Set 1, obtained at the 8th iteration. Now many values of E_{int} are more negative than −70 kcal/mol, including E2 at −73.23 kcal/mol. Values of ΔE_L are all below 10 kcal/mol and values of ΔE_R are mostly below 60 kcal/mol.

Coefficients from the final MLR fit to the converged data in Table 4 gave

Table 4 Iteration No. 8: Docked poses were re-ordered using the SF8 Scoring Function

Ligand	log RBA (expt.)	E_{int}	ΔE_L	ΔE_R	ΔG_{solv}	SF8
E2	2.00	−73.23	3.22	47.88	−12.40	1.92
1a	0.17	−70.99	5.63	40.63	−10.69	−0.04
1b	0.02	−71.47	8.08	50.89	−8.09	−0.53
1c	1.44	−72.26	4.68	38.08	−10.07	0.76
1e	0.45	−72.87	6.15	43.92	−9.58	0.55
1f	1.95	−73.87	4.30	52.38	−8.20	1.93
1 h	−2.40	−65.29	7.82	65.32	−5.97	−2.23
1 k	−0.73	−69.97	5.45	55.36	−7.43	0.15
1 l	0.24	−69.55	4.19	49.39	−8.96	0.31
1o	−1.19	−66.51	7.09	62.51	−8.59	−1.62
1p	−2.10	−62.68	3.15	34.93	−11.15	−2.19
2a	−1.21	−71.28	8.66	44.48	−11.75	−1.11
2e	−2.22	−66.88	8.38	52.66	−8.70	−2.29
3e	−1.42	−69.84	6.27	36.46	−12.32	−0.87
3f	2.28	−75.24	4.30	40.58	−8.98	2.10
4b	1.78	−76.82	5.42	34.71	−9.45	2.04
4e	−0.36	−71.11	6.01	39.85	−11.67	−0.19

$$SF8 = -24.23 - 0.362(E_{int}) - 0.413(\Delta E_L) + 0.0266(\Delta E_R) + 0.0268(\Delta G_{solv}),$$

with $r^2 = 0.942$, rms. dev. = 0.365 log unit and $q^2 = 0.902$. Figure 6a shows that the final correlation with experimental values is indeed excellent, with improved scatter and no outliers. Figure 6b shows that the binding mode for E2 is now consistent with the optimal crystallographic target structure shown in Fig. 2a.

Differences in ligand binding

It is of interest to try to use the previous data in an attempt to understand the differences in binding between strongly

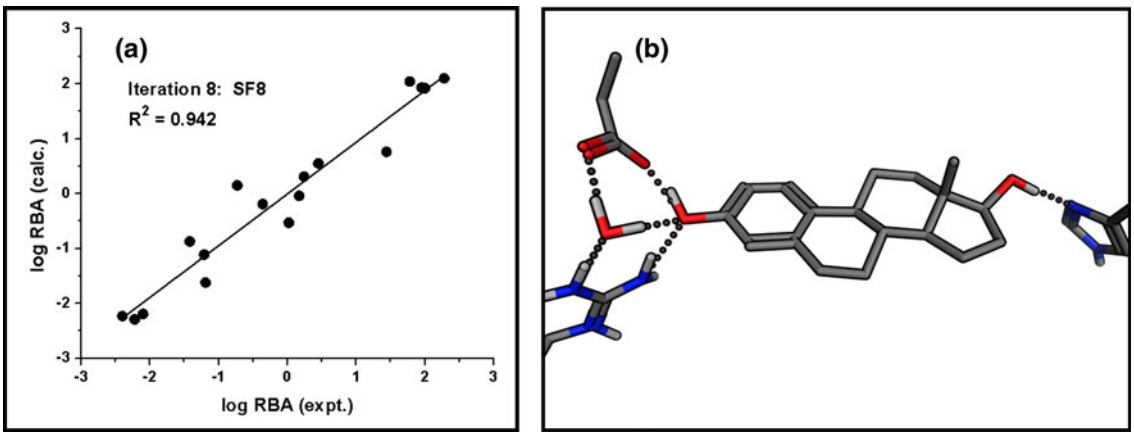

Fig. 6 **a** Correlation between the final iteratively optimized scoring function (SF8) and the experimental log RBA values. **b** Top-ranked binding mode of estradiol, as selected by SF8

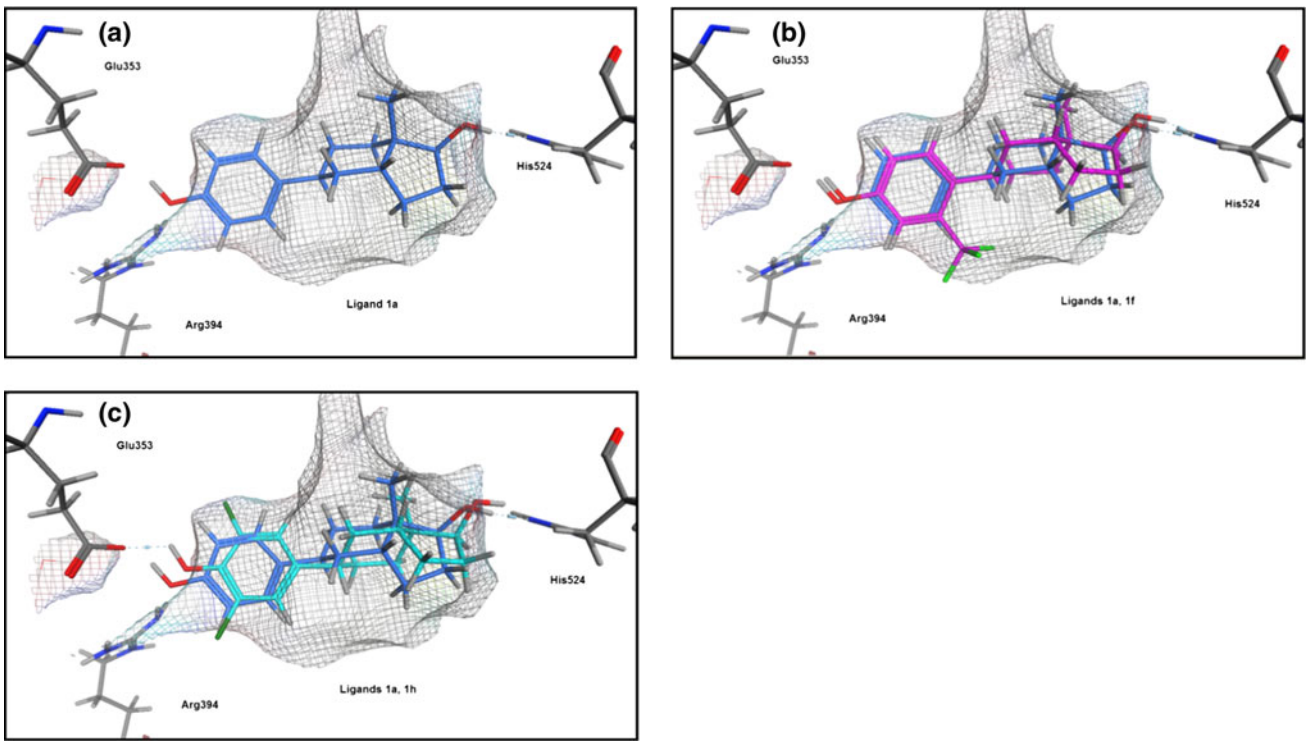

Fig. 7 **a** Parent compound 1a inside a van der Waals suface defining the active site boundary between ligand and receptor. **b** High-RBA compound 1f overlaid over compound 1a **c** Low-RBA compound 1 h overlaid over compound 1a

and weakly bound ligands. For this purpose we chose 1f (5-CF_3 A-CD), 1a, (parent A-CD compound) and 1 h (2,4-diCl A-CD) with experimental log RBAs of 1.95, 0.17, and −2.40, respectively. The computed log RBAs (Table 4, SF8) match the experimental data reasonably well (1.93, −0.04, and −2.23, respectively). This represents roughly equal sampling of log RBA, from the upper end of the range (ca. 2.0) to the lower end of the range (ca. −2.0). The orientation of these three ligands are shown in the van der Waals surface defining the active site boundary between ligand and receptor (Fig. 7). Figure 7a shows the parent ligand 1a, 7b shows compound 1f overlaid over 1a, and 7c shows 1 h overlaid over 1a (the H-atoms play an important role and need to be shown in the active site, but this makes a multiple overlay figure messy and hard to interpret). It can be seen that there are visual differences,

Table 5 Test Set 1: Coefficients taken from converged values for Training Set 1 (SF8)

Ligand	log RBA (expt.)	E_{int}	ΔE_L	ΔE_R	ΔG_{solv}	log RBA (calc.)
1d	1.72	−74.19	5.03	40.30	−9.18	1.40
1 g	−1.40	−66.60	4.99	56.57	−8.53	−0.88
1i	0.66	−73.81	7.27	50.32	−7.13	0.66
1j	−0.42	−69.50	6.36	43.83	−7.25	−0.70
1q	−1.06	−68.43	4.81	50.04	−9.19	−0.33
2f	−2.40	−66.49	10.25	67.16	−6.86	−2.77
3a	−0.61	−71.01	4.32	30.11	−11.77	0.20
3b	2.29	−71.73	3.51	45.69	−10.19	1.25
4a	−0.38	−69.90	4.03	46.48	−11.22	0.37
4f	2.09	−77.54	4.62	41.20	−8.32	2.83

All energy values in kcal/mol

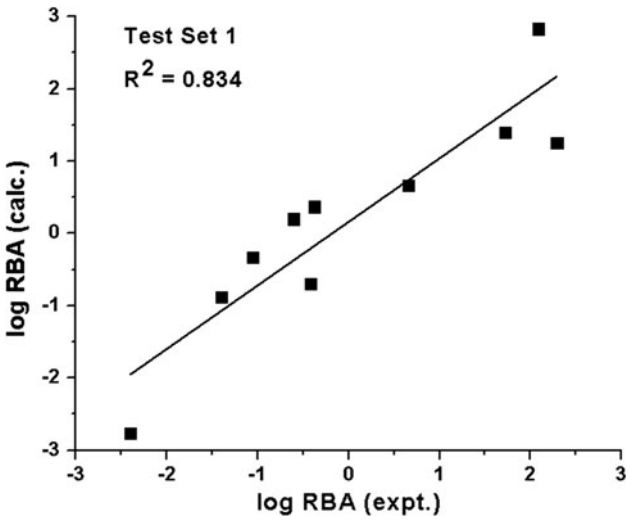

Fig. 8 Calculated versus experimental log RBAs for Test Set 1, using converged coefficients from Training

e.g. the Cl atoms in 1 h push through the surface, crowding nearby residues. This is also shown by examination of Table 4. In the MLR fit for SF8, the weights of the variables in the MLR are: (E_{int}, 1.00), (ΔE_L, 0.55), (ΔE_R, 0.19) and (ΔG_{solv}, 0.04). From Table 4, the magnitude of the interaction energy (more negative is stabilizing) is in the order 1f > 1a ≫ 1 h, and the ligand distortion (less is better) is in the order 1f < 1a ≪ 1 h; the receptor distortion and the solvation energy are relatively unimportant variables. Ignoring receptor distortion and solvation energy, then, from these arguments the scoring function will be in the observed order, i.e. 1f > 1a > 1 h.

Test set

The coefficients obtained from the final converged scoring function SF8 for Training Set 1 were used, unchanged, to rank order and score the docked poses of the Test Set of ligands, i.e. where $C_0 = -24.23$, $C_1 = -0.362$, etc. The docked pose for each ligand with the highest SF8 score was selected, as before, and the calculated binding affinity was compared to the experimental binding affinity. The results are shown in Table 5 and Fig. 8.

Table 5 shows that the range of values for E_{int}, the most important energy descriptor, is from −77.5 to −66.5 kcal/mol for Test Set 1, versus −76.8 to −62.7 kcal/mol for Training Set 1, i.e. the range of values is similar for both. Compound 2f in Test Set 1 appears to have a highly distorted ligand ($\Delta E_L > 10$ kcal/mol) and a highly distorted receptor ($\Delta E_R = 67.2$ kcal/mol). These unfavorable energy contributions are consistent with the fact that the predicted log RBA shows very weak binding (predicted log RBA = −2.77 vs. −2.40 expt.)

Figure 8 shows that Test Set 1, which was not scored by an iterative process, nevertheless shows a very respectable correlation with $r^2 = 0.834$. This implies that the iterative scoring method has genuine predictive value. However, several authors warn of the danger of overfitting inherent in using an iterative method to fit to experimental data (see [30] and references therein) and the whole procedure should be subjected to validation tests to see whether the proposed method is robust and whether it is susceptible to fitting noise instead of actual data. There are two simple ways to do such a test: the first is to generate alternative training and test sets and look for consistency in the accuracy of predicting the experimental RBA values. The second is to randomly scramble the experimental RBA data, associating a randomly chosen RBA for each ligand from the whole set of RBAs (e.g. swap 1a for 2e, 1 g for 3f, etc.). Both methods will be presented here.

Generation of alternate training and test sets

Training Sets 2, 3 and 4 were generated using the same selection conditions described above for Training Set 1. Therefore each set has 17 ligands with representation from high, medium and low RBA values, and from compounds belonging to series 1, 2, 3 and 4. Also, E2 was included in each training set. Each Training Set was fitted using the same iterative method as before.

Figure 9 shows that the convergence properties for the three training sets is similar, with convergence being reached in 5–7 iterations. The results are compared to Training Set 1 in Table 6, and it can be seen that the same pattern is observed, i.e. correlation coefficients based on London dG scoring are initially very low, rise rapidly after 1 iteration, and reach smooth convergence in the range of 0.9 or above. Again, the data are almost perfectly fitted by an exponential function.

Table 6 shows that the four training sets show very good consistency, with values of the correlation coefficient r^2 in

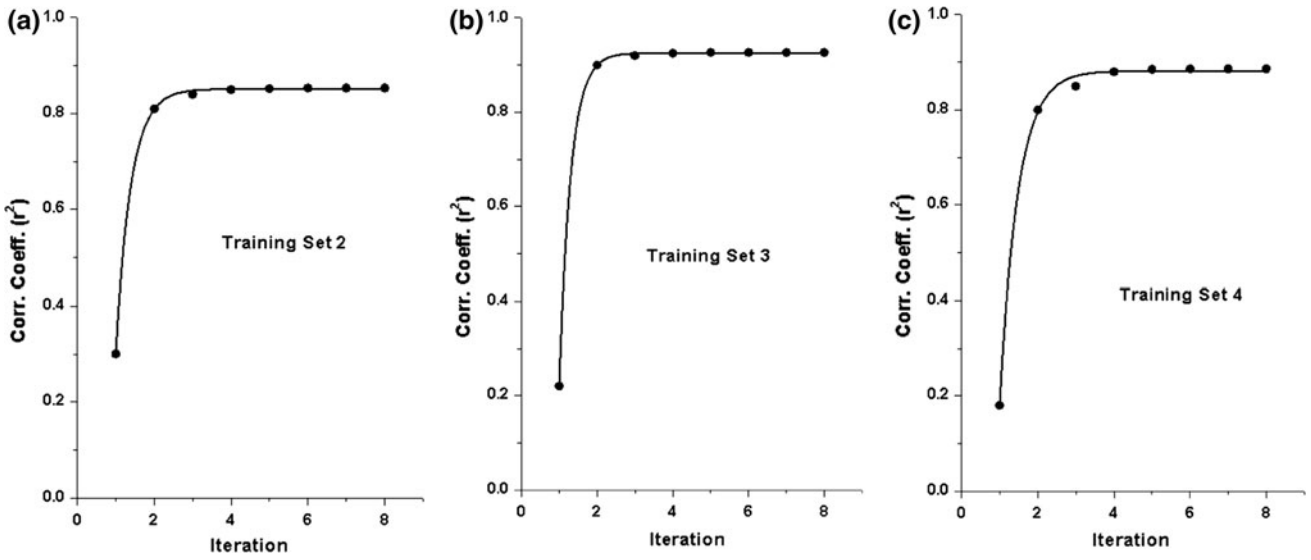

Fig. 9 Convergence properties for new training sets 2, 3, and 4, showing correlation coefficient r^2 versus iteration number. **a** Training Set 2; **b** Training Set 3; **c** Training Set 4

a narrow range of 0.91 ± 0.03. Each of the test sets (10 ligands) formed as the complement to the corresponding training set is also shown in Table 6. Using the converged coefficients from their respective training sets, correlation coefficients r^2 for Test Sets 2, 3 and 4 gave 0.786, .0.859 and 0.927 versus 0.834 for Test Set 1. Thus the iterative method gave an r^2 value of 0.85 ± 0.1 for the four test sets. These results from the four different training and test sets show that the predictions from this iterative procedure are both robust (relatively insensitive to exact composition of training and test set) and of significant predictive value (r^2 and $q^2 > 0.7$ from test sets).

Table 6 Properties of the new Training and Test Sets 2, 3 and 4 are shown along with the original sets 1 for comparison

Training set	Initial r^2	Final r^2	Number of iterations	
1	0.245	0.942	8	
2	0.300	0.900	6	
3	0.219	0.925	3	
4	0.181	0.883	6	
Test set	Composition		r^2	
1	**1d, 1 g, 1i, 1j, 1q, 2f, 3a, 3b, 4a, 4f**		0.834	none
2	**1c, 1f, 1 h, 1 l, 1o, 2a, 3e, 3f, 4b, 4e**		0.786	none
3	**1b, 1e, 1f, 1 l, 1o, 2e, 3b, 3e, 4a, 4b**		0.859	none
4	**1b, 1c, 1f, 1 h, 1p, 2a, 3a, 3f, 4e, 4f**		0.927	none

Scrambling the experimental data for Training Set 1

As a second validation test, the experimental log RBAs were rearranged (scrambled) in a random order, as described above. The log RBA values for the resulting scrambled dataset were completely uncorrelated with the initial experimental log RBA dataset, as shown in Fig. 10, below.

Using these scrambled log RBAs, the same iterative procedure was carried out as described for the Training Sets 1–4 above. Figure 11 shows the variation in correlation coefficient as a function of iteration number for the scrambled data set. Clearly, this time the iterative procedure did not reach smooth convergence but was initially subject to wild oscillations and eventually after 30 iterations reached a repeating oscillatory cycle with a final value of the correlation coefficient below 0.2. This proves that the iterative scoring procedure is not simply overfitting the data, but is progressing towards an optimized set of docked poses when the correct order of experimental log RBAs is used.

Sensitivity of the iterative method to initial choice of DPs

In the iterative method described above, we chose London dG as the original Scoring Function and showed that a very poor correlation with RBA values resulted for the Training Set 1 ($r^2 = 0.245$) If you choose a different (perhaps much better) starting point, will the iterative method result in the same end point? From the published work of Martin using

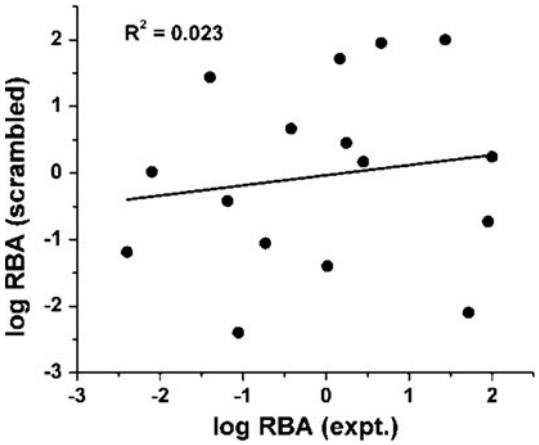

Fig. 10 Correlation between scrambled log RBAs and experimental log RBAs, showing the absence of any significant correlation

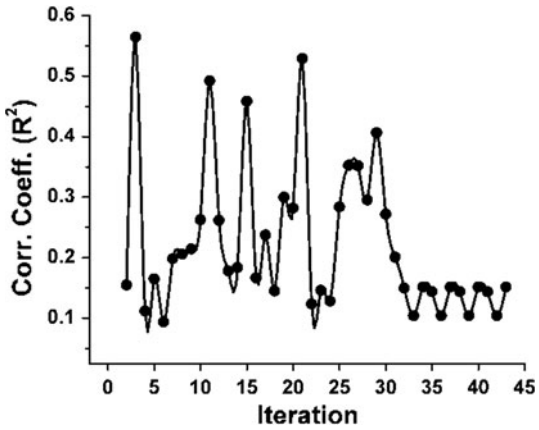

Fig. 11 Value of the Pearson correlation coefficient (r^2) as a function of iteration number for the Training Set 1 with scrambled log RBAs

AutoShim we expect that it will [30]. Observing Tables 2, 3 4 shows that there is in general a trend towards more negative interaction energies as the iterative pose selection proceeds. This suggest that a different and perhaps better starting point to select the best DP would simply be that with the lowest (most negative) E_{int}. In fact, in our previous work [34] we did the pose selection based only on E_{int}, with no iterations. So to address this question we restarted the iterative procedure using the Training Set with the full set of 16,172 DPs and did the first pose selection based on E_{int} rather than on London dG. This gave the following results, presented in Table 7.

Further cycles result in a very small oscillation repeating cycles 6 and 7, so the calculation is converged to within 0.002 in r^2 on continued iteration. Using iteration no. 7 (best value of q^2), the values of r^2, q^2 and rms dev. are 0.942, 0.902 and 0.365, respectively, and the scoring function is $SF7 = -24.23–0.362E_{int}–0.413\Delta E_L + 0.0266\Delta E_R + 0.0268\Delta G_{solv}$. Comparing to the converged parameters and scoring function

Table 7 Correlation parameters for the Training Set 1 of 17 ligands where the initial choice of best DP is that with the lowest value of E_{int}. Definitions of r^2, q^2 and rms dev. are the same as used previously

Iteration Number	r^2	q^2	rms dev. (log units)
1	0.930	0.859	0.402
2	0.927	0.852	0.410
3	0.936	0.852	0.384
4	0.939	0.884	0.375
5	0.942	0.900	0.366
6	0.944	0.893	0.360
7	0.942	0.902	0.365

given below Table 4, where the iterations began with the London dG default value, the parameters (r^2, q^2 and rms dev.) and the derived Scoring Function SF8 are identical to those obtained from SF7 above.

This result is quite remarkable when one considers that the iterative process described previously began with a very poor choice of best docked pose (London dG) whereas the iterative process above began with a very good choice of best docked pose (E_{int}). Converging to the same final value is no coincidence, but rather is a result of obtaining an identical set of 17 best docked poses for the 17 ligands. Indeed one can track the difference in best DPs between the two starting points as a function of iteration number; after 2 iterations there are 4 different DPs used for the MLR, after three iterations there are only two different DPS, after four iterations there is one different DP, and after five iterations the set of DPs is identical. Considering that the Training Set 1 contains a total of 16,172 DPs this is quite a remarkable result, so that the term "robust" for the iterative methodology seems appropriate. It is indeed possible and was suggested by Martin and Sullivan [30] that *any* choice of starting DP for each ligand, e.g. a totally random selection, and the iterative process described in the paper would result in the same choice of best DPs. In a multivariable regression this implies that there is a deep overall minimum in the Scoring Function and that iteration does not trap the Scoring Function in the many local minima which would be expected to exist.

Conclusions

The computational approach to drug discovery is faced with two seemingly incompatible choices: the need for a rapid virtual screening capable of determining whether a compound is likely to be active or inactive, and the prediction of compounds suitable for lead optimization, which requires the prediction of relative binding affinities for ligand to protein. General experience is that the former is

relatively successful, but the latter is much more difficult and currently undergoing continuous development. In this paper we focused on the prediction of relative binding affinities for a set of A-CD estrogens which bind to the receptor ERα, and for which we have determined experimental RBAs.

Our docking procedure is a locally developed modification of the MOE docking software called uDock. It makes extensive use of constraints (tethering) to prepare the receptor so that it stays close to the crystal geometry, while at the same time allowing for relaxation of both the ligand (in this case 17β-estradiol) and the receptor. Once the receptor was prepared the E2 ligand was removed and 26 different A-CD ligands were docked into the active site. Successive relaxation of both ligand and receptor was performed to give a final fully relaxed Complex geometry. Conformer generation (using a cutoff of 7 kcal/mol above the ligand ground state), systematic variation of rotatable bonds, random translation/rotation inside the docking box and energy filtering during docking led to about ca. 500–2,000 acceptable docked poses per ligand.

The next step, creation of a suitable scoring function, is of relevance to lead optimization. From the 27 data values we created a Training Set 1 containing 17 ligands and a Test Set 1 containing 10 ligands. We showed that when the poses are rank-ordered using the default SF in MOE (London dG), and then transformed to an energy-descriptor basis, this Scoring Function for Training Set 1 gave very poor results. Analysis of the top-ranked DPs for each ligand shows that this SF gives a very poor H-bonding network in the active site. However, other lower-ranked DPs gave much better H-bond networks which led to the iterative idea: Why not use the SF just determined to generate a new ranking order for the docked poses? We then continued to re-rank the poses based on the new SF, repeated the MLR, and derived a new correlation. At the second iteration the r^2 value improved dramatically, and the top DPs now showed improved hydrogen bonding. Four more repetitions of this iterative process resulted in a converged r^2 value of 0.942. The DPs at convergence for each ligand were now in accord with experimental data, where known for E2 (i.e. showing the optimum H-bond network, including the bridging water present in the 1GWR pdb file), and may well be in accord with X-Ray structures for the other ligands, if they become available. It is important to note that no additional docking calculations were required in this whole iterative procedure which involved only sorting operations on the DP data file, and sorting is very fast.

The virtues of the entire procedure can now be summarized as follows: (1) once the set of docked poses is obtained, the iterative rescoring procedure is very fast, (2) the rescoring procedure appears to be self-correcting in that the choice of docked poses as measured by the known structure of E2. improves with each iteration, (3) the rescoring and selection of top-ranked docked poses according to maximum score generally improves the correlation with experimental RBAs on each iteration, (4) the whole iterative process converges exponentially to its final value, and (5) standard measures of validity of the data fits show that the fits are robust, even when starting guesses for top-ranked poses are very poor. This latter point was proven by restarting the whole procedure from a very different starting point and reaching an identical end point.

Because of the limited size of the training and test sets, this iterative approach to deriving a scoring function will require further validation before it can be recommended for general use. Our results are consistent, however, with previous results on a variety of other systems obtained by Martin and Sullivan [30] and Jain and coworkers [22–24] using a similar methodology but a far more complex set of descriptors. As described here, the methodology requires use of the uDock program which is linked to the MOE operating environment and a set of only four energy-related descriptors: E_{int}, ΔE_L, ΔE_R and ΔG_{solv}. To avoid the use of different software for calculation of ΔG_{solv} we are currently exploring use of another descriptor relevant to the free energy of solvation already available in MOE, and the results are promising. In general and combined with the experience of other groups, it appears that there is nothing unique to our own choice of algorithms and we expect that a similar approach will have general applicability, given a reasonable set of DPs as input. We hope that with the explicit description given here and in the On-line Information this iterative method of rescoring docked poses will receive the attention it deserves.

Computational methods

Preparation of receptor

Methodology used in this paper has been already been described briefly in ref. 34; a more detailed description is given here. X-Ray crystallographic data for E2 in human recombinant ERα were obtained from the Protein Data Bank as 1GWR. 1GWR is a dimeric protein containing similar chains A and B, with ligands E2 in each active site, coactivator peptides, and associated water molecules. Only dimer B was retained along with its bound ligand E2 and associated water molecules. Here one critical water molecule (HOH2009) is known to form an H-bonded bridge in the active site between Arg394, Glu353 and E2 [34]. Hydrogen atoms were added and partial charges assigned to charged residues at pH 7. Finally, polar H-atoms participating in the H-bond network between E2 (OH on C3), Glu353 and Arg394 and also between E2 (OH on C17) and His524 were adjusted manually to give a near-optimum H-bond network before minimization.

To energy-minimize the E2-HOH-monomer B complex, we used our in-house program H_PDB_Thaw, which is part of the uDock program suite written by Dr. Hooman Shadnia [1]Which uses MOE's Scientific Vector Language, SVL. Details and program listings are available and updated at www.shadnia.com. For all energy minimization calculations we used the MMFF94s force field in the gas phase [37]. To accelerate the calculations, and also for purposes of docking a set of ligands, a 'shell' model of the active site was generated from the full protein (in this case, monomer B). All residues having at least one atom closer than 6 Å to the reference ligand E2 were selected as the inner shell S1. The second shell, S2, was defined as all atoms of all residues having at least one atom closer than 12 Å from the ligand, excluding those already defined in S1. The rest of the atoms were deleted and the broken bonds were capped with hydrogen atoms, making a third shell, S3. Thus, for E2 in ERα, S1 included all the active site atoms, containing 45 residues (664 atoms), S2 included 84 residues (1295 atoms) and S3 contained 25 atoms. The full monomer B contains ca. 4,000 atoms, including hydrogens, so roughly half (2028) of its atoms have been retained.

In order to keep the receptor protein from distorting excessively, the ligand and active site atoms (Shell S1) were initially kept very rigid using strong tether restraints, which are artificial parabolic potentials of the form $k\Delta x_i^2$, where Δx_i is the displacement from the initial atomic coordinate x_i and k is the tether constant. The restraints were then gradually reduced in stages as energy minimization progressed. At the final step all artificial restraints are removed and an energy minimization was performed using the unmodified MMFF94 s force field (for an exact description of the cycles and tether constants, see the H_PDB_Thaw listing, ref. 37). Finally, the ligand E2 was removed and the resulting structure is called the "ERα Prepared Receptor" (ERα-PR). Thus, the thawing algorithm minimizes the movement of active site atoms due to energy minimization, keeping the active site close to the original crystal coordinates.

Docking

To dock the structures in the Training Set plus additional structures in the Test Set into the ERα-PR, the following approach was used. First, a database of ligand conformers was generated using the systematic conformer search utility in MOE. Rotatable bonds were incremented in 20 degree steps and energy-minimized, after which redundant structures were removed. All final conformers within 7 kcal/mol of the lowest-energy conformer were retained for docking.

To generate initial poses, the ligand conformers were randomly placed inside a virtual box called the Docking Box, using a margin of ca. 1 Å larger than the receptor cavity (defined by the van der Waals surface of the active site). The initial complex was then energy-minimized, while the S2 shell atoms were tethered and the S3 shell atoms were fixed but the active shell atoms (S1) were left unconstrained. A docking job was considered to be converged when at least three of the top-ranked docked poses for each ligand were similar (as judged by agreement in total energy and internal energy of the ligand-receptor complex).

Synthetic methods

Synthetic procedures and spectral characterization have been reported previously [33, 34] for compounds in Table 1 belonging to the **1**-series. New RBA data in this paper are given for the compounds in series **2, 3** and **4**. Details of the synthesis of the new structures will be published separately.

Binding affinity

Binding affinities were determined using a competitive radiometric procedure as described in ref. 34. Using this procedure the absolute binding affinity of the reference compound E2 was determined to be 0.2 nM for ERα and 0.5 nM for ERβ. Binding affinities for the other ligands were determined relative to these values.

Acknowledgments This work was funded by the Canadian Breast Cancer Foundation (Ontario Region) grants to J. S. Wright and T. Durst. We are grateful to the CBCF for support for this work.

References

1. Warren GL, Andrews CW, Capelli A-M, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS (2006) A critical assessment of docking programs and scoring functions. J Med Chem 49:5912–5931
2. Hecht D, Fogel GB (2009) Computational intelligence methods for docking scores. Curr Comp-Aided Drug Des 5:56–68
3. Englebienne P, Moitessier N (2009) Docking ligands into flexible and solvated macromolecules. 4. Are popular scoring functions accurate for this class of proteins? J Chem Inf Model 49:1568–1580
4. Englebienne P, Moitessier N (2009) Docking ligands into flexible and solvated macromolecules. 5. Force-field-based prediction of binding affinities of ligands to proteins. J Chem Inf Model 49:2564–2571
5. Moitessier N, Englebienne P, Lee D, Lawandi J, Corbeil CR (2008) Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. Br J Pharmacol 153:S7–S26
6. Corbeil CR, Moitessier N (2009) Docking ligands into flexible and solvated macromolecules. 3. Impact of input ligand conformation, protein flexibility, and water molecules on the accuracy of docking programs. J Chem Inf Model 49:997–1009

[1] uDock and auxiliary programs are available free of charge from Dr. Hooman Shadnia's website at www.shadnia.com (2013).

7. Yoon S, Welsh WJ (2004) Identification of a minimal subset of receptor conformations for improved multiple conformation docking and two-step scoring. J Chem Inf Comput Sci 44:88–96
8. Stjernschantz E, Oostenbrink C (2010) Improved ligand-protein binding affinity predictions using multiple binding modes. Biophys J 98:2682–2691
9. Labute P, Williams C (2001) Flexible alignment of small molecules. J Med Chem 44:1483–1490
10. Chan SL, Labute P (2010) Training a scoring function for the alignment of small molecules. J Chem Inf Model 50:1724–1735
11. Schwarzl SM, Tschopp TB, Smith JC, Fischer S (2002) Can the calculation of ligand binding free energies be improved with continuum solvent electrostatics and an ideal gas entropy correction? J Comput Chem 23:1143–1149
12. Barone V, Cossi M, Tomasi J (1997) A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. J Chem Phys 107:3210–3221
13. Fornabaio M, Cozzini P, Mozzarelli A, Abraham DJ, Kellogg GE (2003) Simple, intuitive calculations of free energy of binding for protein–ligand complexes. 2. Computational titration and pH effects in molecular models of neuraminidase—inhibitor complexes. J Med Chem 46:4487–4500
14. Labute P (2009) Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. Proteins: Struct Funct Bioinf 75:187–205
15. Charifson PS, Corkery JJ, Murcko MA, Walters WP (1999) Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. J Med Chem 42:5100–5109
16. Yang JM, Chen YF, Shen TW, Kristal BS, Hsu DF (2005) Consensus scoring criteria for improving enrichment in virtual screening. J Chem Inf Model 45:1134–1146
17. Lee FS, Chu Z-T, Bolger MB, Warshel A (1992) Calculations of antibody-antigen interactions: microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to McPC603. Protein Eng 5:215–228
18. Aqvist J, Medina C, Samuelsson J-E (1994) A new method for predicting binding affinity in computer-aided drug design. Protein Eng 7:385–391
19. Hansson T, Aqvist J (1995) Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. Protein Eng 8:1137–1144
20. Aqvist J, Hansson T (1996) On the validity of electrostatic linear response in polar solvents. J Phys Chem 100:9512–9521
21. Huang S-Y, Zou X (2006) An iterative knowledge-based scoring function to predict protein- ligand interactions: I. Derivation of the interaction potentials. J Comp Chem 27:1866–1874
22. Jain AN (2006) Scoring functions for protein-ligand docking. Curr Protein Pept Sci 7:407–420
23. Jain AN (1996) Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities. J Comput-Aided Mol Design 10:427–440
24. Jain AN, Harris NL (1995) Park JY quantitative binding site model generation: compass applied to multiple chemotypes targeting the 5-HT1A receptor. J Med Chem 38:1295–1308
25. Durrant JD, McCammon JA (2010) NNscore: a neural-network-based scoring function for the characterization of protein-ligand complexes. J Chem Inf Model 50:1865–1871
26. Durrant JD, McCammon JA (2011) NNScore 2.0: a neural network receptor-ligand scoring function. J Chem Inf Model 51:2897–2903
27. Deng W, Breneman C, Embrechts MJ (2004) Predicting protein-ligand binding affinities using novel geometrical descriptors and machine-learning methods. J Chem Inf Comput Sci 44:699–703
28. Ballester PJ, Mitchell JB (2010) A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. Bioinformatics 26:1169–1175
29. Kinnings SL, Liu N, Tonge PJ, Jackson RM, Xie L, Bourne PE (2011) A machine learning based method to improve docking scoring functions and its application to drug repurposing. J Chem Inf Model 51:408–419
30. Martin E, Sullivan DC (2008) AutoShim: empirically corrected scoring functions for quantitative docking with a crystal structure and IC50 training data. J Chem Inf Model 48:861–872
31. Martin E, Sullivan DC (2008) Surrogate autoshim: predocking into a universal ensemble kinase receptor for three dimensional activity prediction, very quickly, without a crystal structure. J Chem Inf Model 48:873–881
32. Molecular Operating Environment (MOE) (2012) 10; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2012
33. Asim M, El-Safiti M, Qian Y, Choueiri C, Salari S, Cheng J, Shadnia H, Bal M, Pratt MAC, Carlson KE, Katzenellenbogen JA, Wright JS, Durst T (2009) Deconstructing estradiol: removal of B-ring generates compounds which are potent and subtype-selective estrogen receptor agonists. Bioorg Med Chem Lett 19:1250–1253; [Erratum to document cited in (2009) Bioorg Med Chem Lett 19: 2605]
34. Wright JS, Shadnia H, Anderson JM, Durst T, Asim M, El-Safiti M, Choueiri C, Pratt MAC, Ruddy SC, Lau R, Carlson KE, Katzenellenbogen JA, O'Brien PJ, Wan L (2011) A-CD Estrogens: i. Substituent effects, hormone potency, and receptor subtype selectivity in a new family of flexible estrogenic compounds. J Med Chem 54:433–448
35. Merz KM (2010) Limits of free energy computation for protein-ligand combinations. J Chem Theory Comput 6:1769–1776
36. Gaussian 03, revision C.02; Gaussian, Inc.: Wallingford, CT, 2003
37. Halgren TA (1996) Merck molecular force field. I. Basis, form, scope, parameterization and performance of MMFF94. J Comput Chem 17:490–517