Bioorganic & Medicinal Chemistry 20 (2012) 408–414

Contents lists available at SciVerse ScienceDirect

Bioorganic & Medicinal Chemistry

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

Prediction of inhibitory activities of Hsp90 inhibitors

Paolo Swuec [b], David J. Barlow [a,*]

[a] Institute of Pharmaceutical Science, King's College London, Franklin Wilkins Building, 150 Stamford Street, London SE1 9NH, United Kingdom
[b] Department of Pharmaceutical Science, University of Padova, Italy

ARTICLE INFO

Article history:
Received 8 August 2011
Revised 21 October 2011
Accepted 24 October 2011
Available online 30 October 2011

Keywords:
Heat shock proteins
Docking
Neural networks
QSAR
Enzyme inhibition

ABSTRACT

Here, we report on the development of a novel methodology to aid in design of Hsp90 inhibitors, using molecular docking combined with artificial neural network (ANN) modelling. Inhibitors are first docked into the ATPase site of the Human Hsp90α crystal structures and the thermodynamic properties of the complexes together with various physical–chemical properties of the ligands are used as input to train a simple feed-forward, back propagation ANN, to predict the inhibitors' pIC_{50}s. For an objective test set of 60 known Hsp90 inhibitors for which there are no crystallographic data available, the trained ANN is shown to give pIC_{50}s accurate to within ±1 log unit, and the predictions are sufficiently good as to allow the majority of the inhibitors to be ranked correctly according to their potency.

1. Introduction

Heat shock proteins (HSPs) represent a major family of molecular chaperone proteins, present in both prokaryotes and eukaryotes.[1,2] They serve a variety of different roles but are generally concerned with maintaining the integrity and correct functioning of a cell's protein machinery.[3] They are intimately involved in protein turnover and, in some cases, assist in the *de novo* folding of nascent polypeptides, and in guiding the translocation of proteins through membranes.

The chaperone Hsp90 plays a fundamental role in the control of cell growth and development, and hence is essential for cell survival. Under physiological conditions, it is found associated with a wide variety of proteins including calmodulin, actin, tubulin, numerous kinases and several sorts of receptors.[4] Most significantly, it plays a crucial role in stabilizing and maintaining the active conformations of steroid hormone receptors.[3,5] In recent years Hsp90 has been considered a promising target for the treatment of a variety of human diseases. Many of the proteins assisted by Hsp90 are regulators of cell growth and so the enzyme presents an attractive target for cancer chemotherapy,[6] with the expectation that tumour-targeted inhibition of the enzyme would lead to an accumulation of incorrectly folded proteins, inducing the ubiquitin–proteasome system to destroy the abnormal proteins and thereby stimulate apoptosis in the cancer cells.[7–9] Recent studies have also shown that Hsp90 may provide a target for treatment of cardiovascular conditions,[10,11] and neurodegenerative diseases such as Alzheimer's disease,[12] Parkinson's disease,[13] and Amyotrophic Lateral Sclerosis.[14]

Most of the inhibitors so far developed against Hsp90 have been shown to effect their inhibition through binding in the ATP/ADP binding site in its 25 kDa N-terminal ATPase domain.[15,16] Early successes were obtained as a result of serendipitous discovery through large-scale phenotypic screening; inhibitors identified by this means include the benzoquinone ansamicine antibiotics geldanamycin, herbimicin B and radicicol.[17] The subsequent determination of Hsp90 crystal structures later afforded the opportunity for structure-based empirical design programmes, ultimately yielding inhibitors based on a purine scaffold,[18–20] and resorcinolic pyrazole compounds.[21]

The use of Hsp90-inhibitor co-crystal structures has also now allowed virtual screening studies, but the hits so far identified by this means have exhibited only modest (micromolar) activity in inhibition of cancer cell proliferation.[22,23] There are various clinical trials in progress to assess the utility of selected (second generation) Hsp90 inhibitors in the treatment of solid tumours and haematological malignancies, but research into new designs of inhibitors is ongoing and active.

Here, we report on the development of a novel methodology to aid in design of Hsp90 inhibitors, using molecular docking combined with artificial neural network (ANN) modelling. Inhibitors are first docked into the ATPase site of the Human Hsp90α crystal structures and the thermodynamic properties of the complexes together with various physical–chemical properties of the ligands are used as input to train a simple feed-forward, back propagation

* Corresponding author.
E-mail address: dave.barlow@kcl.ac.uk (D.J. Barlow).

0968-0896/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2011.10.069

ANN, to predict the inhibitors' pIC_{50}s. For an objective test set of 60 known Hsp90 inhibitors for which there are no crystallographic data available, the trained ANN is shown to give pIC_{50}s accurate to within ±1 log unit, and the predictions are sufficiently good as to allow the majority of the inhibitors to be ranked correctly according to their potency.

2. Methods

Atomic co-ordinates for Hsp90–inhibitor complexes were downloaded from the Protein Data Bank (PDB, http://www.pdb.org/),[24] with entries initially selected on the criteria that the structures were solved by means of X-ray crystallography and involved only the N-terminal domain (Hsp90α) from *Homo sapiens*. Further restrictions of the dataset were then imposed by selecting only those complexes where data were available on the inhibitor activity expressed as the half maximal inhibitory concentration (IC_{50}) determined using either the Malachite green assay for ATPase activity,[25] or the fluorescence polarisation (FP) assay for Hsp90 inhibitors.[26] The inhibitor activity data were obtained from the Binding Database (BDB, http://www.bindingdb.org/).[27–29]

In order to permit an objective analysis of the prediction methodology, a further 60 ligands were selected, this objective test set comprising inhibitors with known IC_{50} but with no crystallographic data available for their complexes with Hsp90α. The required structural data for complexes involving these inhibitors were generated by means of docking using the software Molegro Virtual Docker (MVD).[30] The IC_{50}s for each inhibitor included in this set were obtained from the studies reported by Brough et al.[31,32]

The Hsp90α–inhibitor complexes selected for use here show a good variation in inhibitor physical–chemical properties. The inhibitors considered include the first discovered antibiotic of the ansamicine class, geldanamycin (GA) (PDB: 1YET),[15] and its derivative 17-desmethoxy-17-*N*,*N*-dimethylaminoethylamino-geldanamycin (17-DMAG) (PDB: 1OSF).[33] There are also 9 crystallographic complexes involving purine-based inhibitor derivatives (PDB: 1UY6; 1UY8; 1UY9; 1UYC; 1UYF; 1UYG; 1UYH; 1UYI; 1UYK).[22] The other ligands taken into account were: pyrazoles with pendant aryl groups (PDB: 2BSM);[34] 3-(5-chloro-2,4-dihydroxyphenyl)-pyrazole-4-carboxamides derivatives (PDB: 2BYH; 2BYI);[35] piperazinyl, morpholino and piperidyl derivatives of pyrazole-based inhibitors (PDB: 2CCS;[36] 2CDD[26]); 8-arylsulfanyl adenines (PDB: 2FWY; 2FWZ; 2H55);[37] 3,4-diarylpyrazole resorcinol pyrazole amide analogues (PDB: 2UWD);[38] 4,5-diarylisoxazole scaffold based structures (PDB: 2VCI; 2VCJ);[32] 2-aminothieno[2,3-*d*]pyrimidine scaffold based structures (PDB: 2WI2; 2WI4; 2WI5; 2WI6; 2WI7);[31] benzisoxazole derivatives (PDB: 3BM9; 3BMY);[39] quinazoline scaffold based structures (PDB: 3FT5; 3FT8);[40] the resorcinol and triazolothione scaffold based structures (PDB: 3HHU).[41]

A Windows© based version of the Pulmonary Drug Delivery Learning engine (PUDDLE)[42] was used for training and testing the artificial neural networks. The ANNs considered used multi layer perceptron network topology and comprised a single input layer, one or more hidden layers, and a single output layer.[43] Networks were trained and tested on a (3.00 GHz; 2.99 GHz) Duo Processor PC equipped with 2.00 Gb RAM.

A set of 10 molecular descriptors was used as ANN inputs. These included properties of the ligands, specifically their volumes, topological polar surface areas (TPSA),[44] and numbers of rotatable bonds (NRTB), together with various computed interaction energies for their complexes with Hsp90α, viz. the H-bonding energy, water–ligand interaction energy, internal ligand interaction energy, external ligand interaction energy, and total energy. They also included the root mean square difference (RMSD) in atomic co-ordinates following least squares superposition of the binding site residues in the complexed and uncomplexed enzyme, and a binary flag to distinguish the assay employed in determination of the inhibitor IC_{50} (for Malachite green assay for ATPase activity:[25] 0; and for fluorescence polarisation assay for Hsp90 inhibitors:[26] 1).

The ligand volume, TPSA, and NRTB were obtained using the web tool Molecular Properties Calculator from Molinspiration Cheminformatics (http://www.molinspiration.com/). The measurement of RMSD was carried out by means of the in-house software, Movement (Barlow, unpublished data). Thermodynamic molecular descriptors were calculated for each Hsp90α–inhibitor complex using Molegro Virtual Docker (MVD).[30] Data were obtained for each complex using the ligand energy inspector tool implemented in MVD, which provides energies computed with both the MolDock and ReRank scoring functions.

All ANN input parameters i were normalised prior to training as:

$$i_n = \frac{i - i_{\min}}{i_{\max} - i_{\min}} \quad (1)$$

where $i_{\min}$ and $i_{\max}$ are the minimum and maximum for each input parameter i, respectively, and i_n is the normalised input.

The docking of the inhibitors in the ATP-binding cavity of the N-terminal Hsp90 domain was carried out with MVD[30] setting a distance constraint to the key aspartic acid residue (Asp93) between 2.30 and 3.60 Å for hydrogen donor atoms in the ligand. The Hsp90α ATPase site was defined using the MVD cavity detection algorithm and the partial charges, bond orders, and flexible torsions for both the protein and the ligand were assigned using the program's preparation tool. The scoring function used was MolDock Score [GRID] with a grid resolution of 0.20 Å. The docking sphere was centred to the detected cavity and the radius set between 8 and 10 Å, depending on the dimensions of the ligand to be docked. MolDock SE was chosen as the search algorithm with the number of runs set to 30. Poses were constrained to the detected cavity and H-bonds optimised after pose prediction. The pose chosen as the best in each case was selected according to the best ReRank Score and the associated interaction energies obtained via the ligand energy inspector.

The 3D structures of the objective test set ligands were built using the Molecular Operative Environment (MOE 2009.10, Chemical Computing Group Inc.) drawing tool, and structure optimisations and partial charge assignments were performed using the MMFF94x force field.

3. Results and discussion

The ANN for use in predicting the pIC_{50}s of Hsp90 inhibitors was trained using the data obtained for a total of 32 inhibitors for which crystal structures were available for their complexes with Human Hsp90α. A comprehensive list of the inhibitors together with their associated molecular descriptors and pIC_{50}s is given in Table 1. The entire set of inhibitors covers a wide range of activity, with IC_{50}s spanning 350 μM to 7 nM. The inhibitors' chemical diversity is such that it includes the four main classes recently reviewed by Janin,[45] with compounds categorised according to their central structural motif and binding mode; in addition to the ansamicine class the set includes resorcinols, (benz)-amides, amino pyr(im)idines, phenols, and purines. For the purposes of ANN training the data set was divided with 26 structures used for network training, and 6 structures used for cross validation (to safeguard against overfitting).

The choice of molecular descriptors to use as ANN input was made so as to cover all possible contributory factors. Given the large gap between the steric effects of geldanamycin (522.1 $Å^3$)

Table 1
The molecular descriptors of each Hsp90 inhibitor used in training and cross validation of the ANN: 10 parameters were selected as inputs, including physico-chemical descriptors such as the ligand molecular volume, and topological polar surface area (TPSA), the protein conformational change induced in Hsp90 upon ligand binding, the inhibitor flexibility, and thermodynamic descriptors

Compound	Vol. (A^3)	TPSA (A^2)[a]	RMSD[b]	NRTB[c]	H-bond En.[d] (kJ mol^{-1})	Wat.–lig. inter.[d] (kJ mol^{-1})	Internal ligand inter.[d] (kJ mol^{-1})	External ligand inter.[d] (kJ mol^{-1})	Tot. energy[d] (kJ mol^{-1})	Assay[e]
1OSF	588.48	169.529	2.409	8	−12.024	−46.631	7.772	−196.398	−188.626	0
3FT5	161.272	51.807	0.22	0	−1.98	−18.163	5.159	−74.111	−68.952	0
2WI2	134.113	64.699	0.297	1	−1.98	−18.848	5.081	−66.106	−61.026	0
2WI5	298.055	141.058	0.411	4	−6.2	7.842	14.516	−37.682	−23.166	0
1YET	522.114	163.498	2.409	5	−8.955	−19.782	20.911	−152.242	−131.332	0
2H55	366.041	100.128	1.411	7	−2.347	−47.962	14.717	−161.979	−147.262	0
1UYF	368.2	97.335	1.362	8	−2.375	−34.493	21.137	−158.352	−137.215	1
2CCS	247.404	84.404	0.204	2	−3.187	−28.22	6.037	−121.761	−115.724	0
2VCJ	393.528	108.058	0.377	6	−6.005	−13.47	23.302	−142.087	−118.786	0
2WI4	276.012	76.535	1.347	3	−3.323	−18.585	12.593	−33.803	−21.21	0
2VCI	429.942	108.058	0.324	7	−6.548	−20.341	21.209	−173.988	−152.778	0
1UYK	296.259	88.101	1.295	5	−2.24	−8.089	12.658	−98.687	−86.029	1
2CDD	314.381	87.607	0.245	3	−3.421	−15.891	15.727	−135.035	−119.308	0
2FWY	356.331	100.128	1.564	7	−2.27	−27.333	17.572	−133.971	−116.399	0
1UYG	256.073	98.958	1.23	4	−2.834	−36.927	19.229	−139.512	−120.283	1
2WI7	396.689	93.377	1.471	7	−4.272	−27.807	16.662	−135.363	−118.701	0
3BMY	330.567	90.987	1.348	5	−3.932	−18.774	21.963	−119.969	−98.006	0
3HHU	378.096	109.506	0.431	7	−4.326	−26.126	17.323	−154.116	−136.794	0
2WI6	284.952	80.905	1.746	3	−3.208	−13.295	6.874	−114.438	−107.565	0
1UY6	344.036	97.335	1.452	8	−2.66	−40.521	22.629	−155.231	−132.602	1
1UY8	292.944	78.867	1.336	6	−2.224	−33.257	19.071	−129.579	−110.508	1
1UYC	318.49	88.101	1.316	7	−2.647	−39.942	19.784	−145.408	−125.624	1
1UYH	323.421	88.101	1.236	7	−2.399	−37.973	17.001	−148.183	−131.183	1
2BYI	328.169	158.402	1.653	5	−7.794	−42.368	13.277	−165.086	−151.809	0
3BM9	218.794	86.716	0.204	1	−6.349	−30.332	8.936	−120.541	−111.606	0
3FT8	318.04	97.291	1.361	2	−2.32	−46.251	14.992	−165.017	−150.025	0
2UWD	324.138	104.82	0.337	5	−5.28	−44.026	10.815	−169.823	−159.008	0
2FWZ	362.436	100.128	1.566	7	−2.632	−38.239	14.97	−144.925	−129.955	0
1UY9	291.328	88.101	1.421	5	−2.248	−42.457	13.674	−144.395	−130.721	1
2BYH	304.192	115.308	1.815	4	−6.797	−41.6	10.692	−151.844	−141.152	0
2BSM	327.556	107.471	0.281	5	−7.212	−24.637	12.59	−152.61	−140.02	0
1UYI	329.119	88.101	1.387	7	−2.8	−40.353	19.359	−157.248	−137.889	1

[a] Topological polar surface area of a molecule defined as the surface sum over all polar atoms, including the surface area of attached hydrogens.
[b] Root mean square difference in the atomic positions for the atoms of proteins residues within 4 Å of the bound inhibitor.
[c] Number of rotable bonds.
[d] Energy descriptors obtained by Molegro Virtual Docker, see Section 2 for details.[30]
[e] Input 1 for Malachite green assay for ATPase activity,[25] input 0 for fluorescence polarisation assay for Hsp90 inhibitors.[26]

(PDB code 1YET) and a small inhibitor like 4-methyl-6-(methylsulfanyl)-1,3,5-trazin-2-amine (134.1 $Å^3$) (PDB code 2WI2), and also given the wide gap between their activities (0.020 μM and 350 μM, respectively), the introduction of molecular volume as input was considered appropriate. In view also of the fact that Hsp90α ligands are shown to require a H-bond donor group in order to mimic the endogen ADP, with other crucial interactions involving the key conserved aspartic acid residue (Asp93), and the network of water molecules at the bottom of the ATPase cavity, it was clearly essential too to have ANN inputs quantifying the protein–ligand H-bonding Energy and polar surface area (TPSA). Since the N-terminal domain of Hsp90 undergoes significant conformational changes when ligands bind,[46] the edge of the hydrophobic pocket being flexible and the accommodation of larger groups made possible by the movement of the loop between Ile104 and Ala111,[22] it was also deemed important to have ANN inputs to quantify the extent of conformational change in the ligand and protein. For the protein, we considered the conformational changes in residues within 4 Å of any ligand atom comparing against the positions of the same residues in the uncomplexed protein (PDB code 1UYL).[22] The results were expressed as the all atom root mean square deviation (RMSD in Å) for the entire set of binding site residues. To quantify ligand flexibility we used an ANN input of the number of rotatable bonds (NRTB).

The topology of the ANN was determined by trial and error, manually optimising the number of hidden neurons, the network momentum, the learning rate, and noise level. For the final trained network we settled on a topology of 10 input nodes, nine hidden nodes, and one output (overall structure 10-9-1). This network was found to give best results, considering the mean square errors in pIC_{50} predictions made for the training, and cross validation data sets.

Considering the pIC_{50} predictions made for the training and cross validation data sets using the crystallographic data as inputs (Table 2), it is clear that the ANN was successfully trained and the problem of overfitting avoided; regression of the predicted on the measured pIC_{50}s gives a strong positive correlation, with a slope and intercept not significantly different from 1 and 0, respectively (Fig. 1). The pIC_{50} predictions made for the same set of 32 ligands, using ANN inputs furnished through MVD re-dockings, gave equally encouraging results (Table 2).

Considering the predictions made for the training and cross validation data sets using re-docking data as inputs, the regression of the predicted on the measured pIC_{50}s again gives a strong positive correlation, with a slope and intercept not significantly different from 1 and 0, respectively (accepting significance as $P_\alpha \leqslant 0.05$).

For the 60 inhibitors in the objective test set the pIC_{50} predictions are not as accurate (Table 3) but the ranking of the predicted pIC_{50}s is more or less correct (giving a Spearman's rank Correlation Coefficient, $R = 0.84$, with $P_\alpha \leqslant 0.01$).

The IC_{50} for the most potent inhibitors are predicted with higher accuracy than those for weak inhibitors, and the general trend is for the ANN to overestimate the potency of the objective test set inhibitors. On average, the predicted IC_{50}s are in error by around

Table 2
The actual and the predicted pIC_{50} of each inhibitor used to train and cross validation of the network. The final ANN configuration adopted is (10-9-1)

Compound	IC_{50}[b] (nM)	Actual pIC_{50}[a]	Predicted pIC_{50} crystal input (Err.)	Predicted pIC_{50} ReDock input (Err.)
1OSF	24	7.62	7.68 (−0.06)	7.58 (−0.04)
3FT5	15,000	4.82	4.87 (−0.04)	5.08 (−0.26)
2WI2	350,000	3.46	3.48 (−0.03)	3.74 (−0.28)
2WI5	900	6.05	6.1 (−0.05)	6.85 (−0.8)
1YET	20	7.7	7.72 (−0.02)	7.45 (0.25)
2H55	80	7.1	7.11 (−0.01)	6.84 (0.26)
1UYF	1500*	5.82	5.82 (0)	5.53 (0.3)
2CCS	2000	5.7	5.78 (−0.08)	6.39 (−0.69)
2VCJ	21	7.68	7.86 (−0.18)	7.65 (0.03)
2WI4	1560	5.81	5.88 (−0.07)	6.47 (−0.67)
2VCI	7	8.15	8.23 (−0.08)	7.98 (0.18)
1UYK	17,100*	4.77	4.81 (−0.04)	5.45 (−0.68)
2CDD	148	6.83	6.78 (0.05)	6.97 (−0.14)
2FWY	205	6.69	6.82 (−0.14)	6.46 (0.23)
1UYG	53,500*	4.27	4.35 (−0.07)	4.61 (−0.34)
2WI7	58	7.24	7.49 (−0.26)	7.44 (−0.2)
3BMY	30	7.52	7.38 (0.15)	6.97 (0.55)
3HHU	41	7.39	7.26 (0.13)	7.97 (−0.59)
2WI6	230	6.64	6.68 (−0.04)	6.74 (−0.11)
1UY6	1000*	6	5.85 (0.15)	5.46 (0.54)
1UY8	75,000*	4.12	4.02 (0.11)	3.66 (0.46)
1UYC	41,000*	4.39	4.89 (−0.5)	4.73 (−0.34)
1UYH	14,300*	4.84	4.75 (0.1)	4.52 (0.32)
2BYI	461	6.34	6.46 (−0.12)	6.05 (0.28)
3BM9	190	6.72	6.64 (0.08)	6.56 (0.16)
3FT8	30	7.52	7.59 (−0.06)	7.62 (−0.1)
Compound	IC_{50}[a] (nM)	Actual pIC_{50}[b]	Predicted pIC_{50} crystal input[c] (Err.)	Predicted pIC_{50} ReDock input[d] (Err.)
2UWD	28	7.55	7.35 (0.2)	7.02 (0.53)
2FWZ	50	7.3	6.99 (0.31)	6.9 (0.41)
1UY9	15,300*	4.82	4.59 (0.23)	4.45 (0.37)
2BYH	258	6.59	7.07 (−0.48)	7.16 (−0.57)
2BSM	25	7.6	7.6 (0)	7.33 (0.27)
1UYI	4100*	5.39	5.55 (−0.16)	6.07 (−0.68)

[a] Values marked with * referred to Malachite green assay for ATPase activity,[25] values not marked with * referred to fluorescence polarisation (FP) assay for Hsp90 inhibitors.[26]
[b] The activity is expressed as a function of the negative logarithm of molar concentration: $pIC_{50} = -\log_{10} IC_{50}$.
[c] Input data from the crystallographic structure of the protein–inhibitor PDB complex.
[d] Input data from the best re-docked pose by MVD.[30]

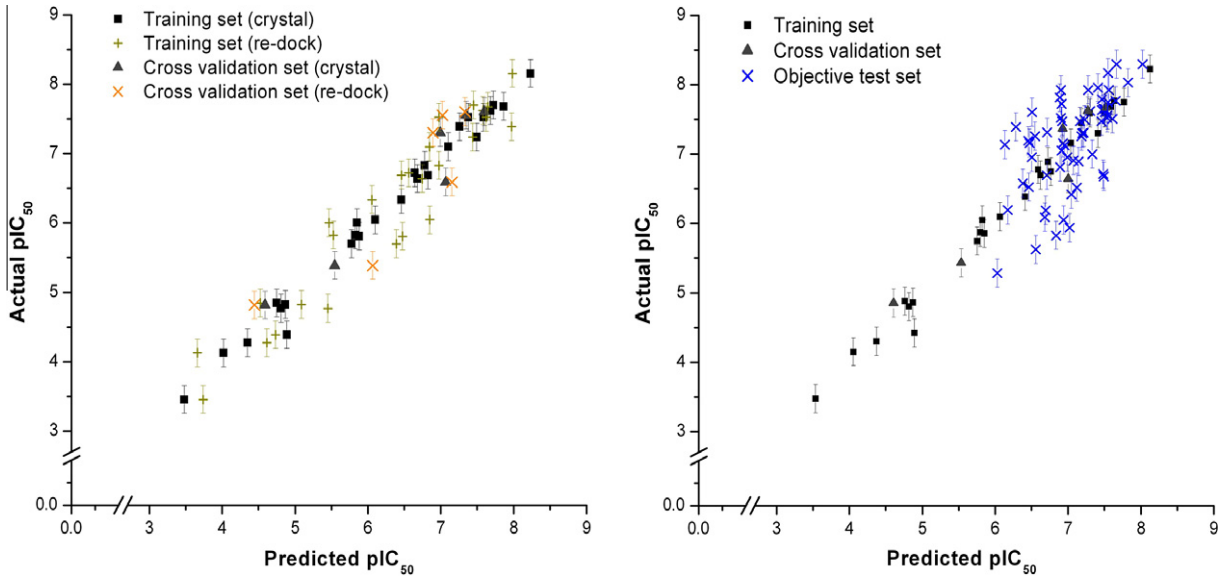

Figure 1. Correlation between the observed pIC_{50}s and the predicted pIC_{50}s for the training, cross validation and objective test data sets.

51% which should be judged in the context of experimental uncertainties that are frequently also of this magnitude, cf., the error of ±47% quoted for inhibitor 4-bromo-6-(6-hydroxy-1,2-benzisoxazole-3-yl)benzene-1,3-diol (PDB: 3BM9).[26,36,39]

Analysis of the ANN connection weights indicates that nine of the ten inputs have similar influence in determining the inhibitor activity. The exception is NRTB (the number of rotatable bonds in the ligand) which is about two thirds as important as the other

Table 3
The actual and the predicted pIC_{50} of each inhibitor of the objective test set. The final ANN configuration adopted is (10-9-1)

Compound	IC_{50} (nM)	Actual pIC_{50}[a]	Predicted pIC_{50} (Err.)	Compound	IC_{50} (nM)	Actual pIC_{50}[a]	Predicted pIC_{50} (Err.)
1	5700	5.24	6.06 (−0.82)	**31**	914	6.04	6.74 (−0.7)
2	720	6.14	6.21 (−0.07)	**32**	1660	5.78	6.9 (−1.12)
3	340	6.47	6.51 (−0.04)	**33**	21	7.68	7.64 (0.04)
4	300	6.52	6.43 (0.1)	**34**	14	7.85	6.97 (0.89)
5	56	7.25	6.77 (0.48)	**35**	39	7.41	7.54 (−0.13)
6	64	7.19	6.6 (0.6)	**36**	64	7.19	7.26 (−0.07)
7	47	7.33	6.32 (1)	**37**	19	7.72	7.55 (0.17)
8	127	6.9	6.55 (0.34)	**38**	28	7.55	7.38 (0.17)
9	80	7.1	7 (0.1)	**39**	39	7.41	7.24 (0.17)
10	29	7.54	6.56 (0.98)	**40**	18	7.74	6.95 (0.79)
11	80	7.1	6.52 (0.58)	**41**	127	6.9	7.06 (−0.16)
12	176	6.75	6.95 (−0.2)	**42**	343	6.46	7.19 (−0.73)
13	87	7.06	7.02 (0.04)	**43**	239	6.62	7.57 (−0.95)
14	102	6.99	6.97 (0.02)	**44**	32	7.49	7.63 (−0.14)
15	84	7.08	6.17 (0.9)	**45**	29	7.54	7.59 (−0.05)
16	58	7.24	7.27 (−0.03)	**46**	13	7.89	7.49 (0.39)
17	146	6.84	7.22 (−0.38)	**47**	20	7.7	7.75 (−0.05)
18	115	6.94	7.41 (−0.47)	**48**	26	7.59	7.59 (0)
19	35	7.46	6.97 (0.49)	**49**	26	7.59	7.57 (0.01)
20	27	7.57	7.55 (0.02)	**50**	22	7.66	6.97 (0.68)
21	37	7.43	7.31 (0.13)	**51**	14	7.85	7.35 (0.5)
22	57	7.24	7.29 (−0.04)	**52**	39	7.41	6.98 (0.43)
23	142	6.85	7.15 (−0.3)	**53**	14	7.85	7.65 (0.2)
24	222	6.65	7.57 (−0.91)	**54**	36	7.44	7.7 (−0.25)
25	1000	6	7(−1)	**55**	35	7.46	7.63 (−0.17)
26	231	6.64	6.77 (−0.13)	**56**	8	8.1	7.64 (0.46)
27	2630	5.58	6.61 (−1.03)	**57**	11	7.96	7.92 (0.04)
28	728	6.14	6.75 (−0.61)	**58**	6	8.22	7.76 (0.47)
29	1290	5.89	7.09 (−1.2)	**59**	6	8.22	8.13 (0.1)
30	431	6.37	7.12 (−0.75)	**60**	74	7.13	6.5 (0.63)

[a] The activity is expressed as a function of the negative logarithm of molar concentration: $pIC_{50} = -\log_{10} IC_{50}$.

inputs. It is also worthy of note here that the type of assay was required as an ANN input, thus confirming the reported difference in sensitivity of the Malachite green and fluorescence polarisation assays.[26]

As widely reported, Hsp90 plasticity is a key factor in its binding of ligands and the necessity to include the All at. RMSD as input is thus not unexpected.[7,46] The poorer quality of the pIC_{50} predictions for the objective test set ligands may be readily attributed therefore, to approximations made as regards the extent of the conformational changes induced by the ligand binding. This could be quantified for the training and cross validation data set inhibitors, by comparing the structures of the complexed and uncomplexed enzymes. For the objective test set inhibitors, however, the required RMSD could be obtained only by allowing MVD freedom to alter the conformations of binding site residues during the docking. With such freedom allowed, however, the results obtained were very poor (and the pIC_{50} predictions correspondingly highly inaccurate). We performed the docking on the Apo structure of Hsp90α (PDB: 1UYL) allowing displacement of water molecules inside the cavity and sidechain conformational changes, by means of MVD displacement water and MVD sidechain flexibility functions, respectively. Even employing these optional algorithms the docking returned unacceptable thermodynamic values, due to high entropy and clash penalty, and poor prediction results with these inputs presented to the ANN. An alternative strategy was thus adopted to dock the objective ligand in the target taken from the crystallographic complex of the inhibitor belonging to the same chemotype class (target from PDB complex 2VCI[31] to dock objective ligands 1–16, and PDB complex 2WI7[32] to dock objective ligands 17–60). This permitted to overcome the poor simulation of protein conformational change taking advantage of the conformational state induced by the crystallized inhibitor. We also performed the docking recruiting the MVD displacement water and MVD sidechain flexibility functions but the ANN returned poor quality results.

A comparison of the prediction quality between input provided by MolDock and ReRank scoring functions was performed. We noticed that thermodynamic descriptors obtained by means of ReRank scoring function led to higher accuracy in pIC_{50} predictions than data given by MolDock scoring function.

Inhibitor 3FT8 gave a predicted pIC_{50} error of (−0.06) when the prediction was made using crystallographic data, and a predicted pIC_{50} error of (−0.10) when the best re-docked pose's thermodynamic data were used as input. Inhibitor 2WI5 gave a predicted pIC_{50} error of (−0.06) when the prediction was made using crystallographic data by means of crystallographic data as input, and a predicted pIC_{50} error of (−0.08) if the prediction was made on the basis of the best re-docked pose data. The investigation of such a variety in prediction quality led to the analysis of poses generated by the re-docking engine. The RMSD from the reference ligand was found to be 0.256 for 3FT8 (Fig. 2a), while the RMSD for 2WI5 was found to be 6.660 from reference ligand (Fig. 2b). This sheds light on the importance of the docking quality, and with the previously made observation on the ineffective sidechain flexibility tool, it is easy to understand that the limit as regards the accuracy of pIC_{50} predictions is not determined by the ANN method itself but by the quality of input data provided through the simulation of ligand docking.

4. Conclusion

Developing new operating schemes as tools for scoring the activities of novel compounds is a major challenge in drug discovery research. Through interaction with multiple substrates, Hsp90 provides a bridge between the cell and its environment, allowing the integration of the stress response to the normal transductional and transcriptional processes. The experimental work detailed here has demonstrated that an artificial neural network of the proper type and configuration can be trained to predict the half maximum inhibitory concentration of Hsp90 inhibitors with

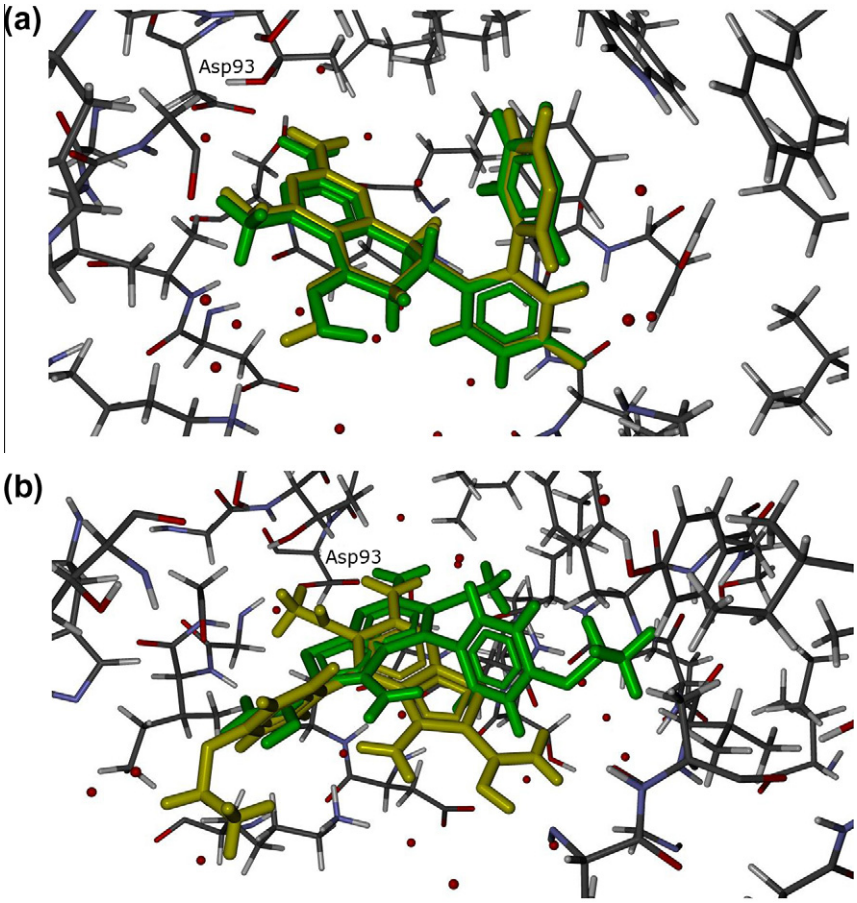

Figure 2. (a) Shows the different orientation in the cavity for the crystallographic structure 3FT8 (green) and for the best re-docked pose of the same ligand according to ReRank Scoring function (yellow); (b) shows the different orientation in the cavity for the crystallographic structure 2WI5 (green) and for the best re-docked pose according to ReRank Scoring function (yellow). The RMSD from the reference ligand was found to be 0.256 for 3FT8 associated to an error in prediction of (−0.05), and 6.660 for 2WI5 associated to an error in prediction of (−0.80).

considerable success. The reliability of the method is limited primarily by the quality of the ANN input data furnished through molecular docking. The model here presented may provide a useful tool for investigating the potential Hsp90 inhibitory activities of novel compounds—of chemotypes represented within the existing (Hsp90) crystallographic data set—in order to accelerate and improve the discovery of new drug leads. Such modelling might be extended to any molecular target for which there is sufficient crystallographic data, and whose inhibitory mechanism can be parameterized as it was for Hsp90.

Acknowledgement

The authors gratefully acknowledge support from European Union within the Erasmus student exchange program.

Supplementary data

Supplementary data (the input data for the objective test set, the input data obtained by means of re-docking for training and cross validation sets, and the SMILE entry for each inhibitor structure included in the objective test set with the activity value) associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2011.10.069. These data include MOL files and InChiKeys of the most important compounds described in this article.

References and notes

1. Ritossa, F. *Cell. Mol. Life Sci.* **1962**, *18*, 571.
2. Tissieres, H. K.; Mitchell, U. M. *J. Mol. Biol.* **1974**, *84*, 389.
3. Csermely, P.; Schnaider, T.; Soti, C.; Prohaszka, Z.; Nardai, G. *Pharmacol. Ther.* **1998**, *79*, 129.
4. Miyata, Y.; Yahara, I. *J. Biol. Chem.* **1992**, *267*, 7042.
5. Buchner, J. *Trends Biochem. Sci.* **1999**, *24*, 136.
6. Didelot, C.; Lanneau, D.; Brunet, M.; Joly, A. L.; De Thonel, A.; Chiosis, G.; Garrido, C. *Curr. Med. Chem.* **2007**, *14*, 2839.
7. Xu, D. M.; Zalmas, L. P.; La Thangue, N. B. *EMBO Rep.* **2008**, *9*, 662.
8. Dai, C.; Whitesell, L.; Rogers, A. B.; Lindquist, S. *Cell* **2007**, *130*, 1005.
9. Santoro, M. G. *Biochem. Pharmacol.* **2000**, *59*, 55.
10. Benjamin, J.; McMillan, D. R. *Circ. Res.* **1998**, *83*, 117.
11. Antonova, G.; Lichtenbeld, H.; Xia, T.; Chatterjee, A.; Dimitropoulou, C.; Catravas, J. D. *Clin. Hemorheol. Microcirc.* **2007**, *37*, 19.
12. Wyttenbach; Patrick, A. A. *The Role of Heat Shock Proteins During Neurodegeneration in Alzheimer's, Parkinson's and Huntington's Disease*; Landes Bioscience, 2007. pp 81–99.
13. Mclean, P. J. In *Cellular and Molecular Mechanisms Underlying Parkinson's Disease: The Role of Molecular Chaperones*; Springer, 2008; Vol. 3, pp 51–68.
14. Kalmar, B.; Greensmith, L. In *Heat Shock Proteins as Therapeutic Targets in Amyotrophic Lateral Sclerosis*; Springer, 2008; Vol. 3, pp 69–107.
15. Stebbins, C. E.; Russo, A. A.; Schneider, C.; Rosen, N.; Hartl, F. U.; Pavletich, N. P. *Cell* **1997**, *89*, 239.
16. Goetz, M. P.; Toft, D. O.; Ames, M. M.; Erlichman, C. *Ann. Oncol.* **2003**, *14*, 1169.
17. Patel, H. J.; Modi, S.; Chiosis, G.; Taldone, T. *Expert Opin. Drug Discov.* **2011**, *6*, 559.
18. Chiosis, G.; Timaul, M. N.; Lucas, B.; Munster, P. N.; Zheng, F. F.; Sepp-Lorenzino, L.; Rosen, N. *Chem. Biol.* **2001**, *8*, 289.
19. Taldone, T.; Chiosis, G. *Curr. Top. Med. Chem.* **2009**, *9*, 1436.
20. Wright, L.; Barril, X.; Dymock, B.; Sheridan, L.; Surgenor, A.; Beswick, M.; Drysdale, M.; Collier, A.; Massey, A.; Davies, N.; Fink, A.; Fromont, C.; Aherne, W.; Boxall, K.; Sharp, S.; Workman, P.; Hubbard, R. E. *Chem. Biol.* **2004**, *11*, 775.

21. Zhou, V.; Han, S. L.; Brinker, A.; Klock, H.; Caldwell, J.; Gu, X. J. *Anal. Biochem.* **2004**, *331*, 349.
22. Park, H.; Kim, Y.-J.; Hahn, J.-S. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 6345.
23. Thorat, D. A.; Doddareddy, M. R.; Seo, S. H.; Hong, T.-J.; Cho, Y. S.; Hahn, J.-S.; Pae, A. N. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 1593.
24. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.
25. Rowlands, M. G.; Newbatt, Y. M.; Prodromou, C.; Pearl, L. H.; Workman, P.; Aherne, W. *Anal. Biochem.* **2004**, *327*, 176.
26. Howes, R.; Barril, X.; Dymock, B. W.; Grant, K.; Northfield, C. J.; Robertson, A. G. S.; Surgenor, A.; Wayne, J.; Wright, L.; James, K.; Matthews, T.; Cheung, K. M.; McDonald, E.; Workman, P.; Drysdale, M. J. *Anal. Biochem.* **2006**, *350*, 202.
27. Liu, T. Q.; Lin, Y. M.; Wen, X.; Jorissen, R. N.; Gilson, M. K. *Nucleic Acids Res.* **2007**, *35*, 198.
28. Chen, X.; Lin, Y. M.; Liu, M.; Gilson, M. K. *Bioinformatics* **2002**, *18*, 130.
29. Chen, X.; Lin, Y.; Gilson, M. K.; Binding, The. *Biopolymers Nucleic Acid Sci.* **2002**, *61*, 127.
30. Thomsen, R.; Christensen, M. H. *J. Med. Chem.* **2006**, *49*, 3315.
31. Brough, P. A.; Barril, X.; Borgognoni, J.; Chene, P.; Davies, N. G. M.; Davis, B.; Drysdale, M. J.; Dymock, B.; Eccles, S. A.; Garcia-Echeverria, C.; Fromont, C.; Hayes, A.; Hubbard, R. E.; Jordan, A. M.; Jensen, M. R.; Massey, A.; Merrett, A.; Padfield, A.; Parsons, R.; Radimerski, T.; Raynaud, F. I.; Robertson, A.; Roughley, S. D.; Schoepfer, J.; Simmonite, H.; Sharp, S. Y.; Surgenor, A.; Valenti, M.; Walls, S.; Webb, P.; Wood, M.; Workman, P.; Wright, L. *J. Med. Chem.* **2009**, *52*, 4794.
32. Brough, P. A.; Aherne, W.; Barril, X.; Borgognoni, J.; Boxall, K.; Cansfield, J. E.; Cheung, K. M. J.; Collins, I.; Davies, N. G. M.; Drysdale, M. J.; Dymock, B.; Eccles, S. A.; Finch, H.; Fink, A.; Hayes, A.; Howes, R.; Hubbard, R. E.; James, K.; Jordan, A. M.; Lockie, A.; Martins, V.; Massey, A.; Matthews, T. P.; McDonald, E.; Northfield, C. J.; Pearl, L. H.; Prodromou, C.; Ray, S.; Raynaud, F. I.; Roughley, S. D.; Sharp, S. Y.; Surgenor, A.; Walmsley, D. L.; Webb, P.; Wood, M.; Workman, P.; Wrightt, L. *J. Med. Chem.* **2008**, *51*, 196.
33. Jez, J. M.; Chen, J. C. H.; Rastelli, G.; Stroud, R. M.; Santi, D. V. *Chem. Biol.* **2003**, *10*, 361.
34. Dymock, B. W.; Barril, X.; Brough, P. A.; Cansfield, J. E.; Massey, A.; McDonald, E.; Hubbard, R. E.; Surgenor, A.; Roughley, S. D.; Webb, P.; Workman, P.; Wright, L.; Drysdale, M. J. *J. Med. Chem.* **2005**, *48*, 4212.
35. Brough, P. A.; Barril, X.; Beswick, M.; Dymock, B. W.; Drysdale, M. J.; Wright, L.; Grant, K.; Massey, A.; Surgenor, A. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 5197.
36. Barril, X.; Beswick, M. C.; Collier, A.; Drysdale, M. J.; Dymock, B. W.; Fink, A.; Grant, K.; Howes, R.; Jordan, A. M.; Massey, A.; Surgenor, A.; Wayne, J.; Workman, P.; Wright, L. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2543.
37. Immormino, R. M.; Kang, Y. L.; Chiosis, G.; Gewirth, D. T. *J. Med. Chem.* **2006**, *49*, 4953.
38. Sharp, S. Y.; Prodromou, C.; Boxall, K.; Powers, M. V.; Holmes, J. L.; Box, G.; Matthews, T. P.; Cheung, K. M. J.; Kalusa, A.; Janmes, K.; Hayes, A.; Hardcastle, A.; Dymock, B.; Brough, P. A.; Barril, X.; Cansfield, J. E.; Wright, L.; Surgenor, A.; Foloppe, N.; Hubbard, R. E.; Aherne, W.; Pearl, L.; Jones, K.; McDonald, E.; Raynaud, F.; Eccles, S.; Drysdale, M.; Workman, P. *Mol. Cancer Ther.* **2007**, *6*, 1198.
39. Gopalsamy; Shi, M. X.; Golas, J.; Vogan, E.; Jacob, J.; Johnson, M.; Lee, F.; Nilakantan, R.; Petersen, R.; Svenson, K.; Chopra, R.; Tam, M. S.; Wen, Y. X.; Ellingboe, J.; Arndt, K.; Boschelli, F. *J. Med. Chem.* **2008**, *51*, 373.
40. Barker, J. J.; Barker, O.; Boggio, R.; Chauhan, V.; Cheng, R. K. Y.; Corden, V.; Courtney, S. M.; Edwards, N.; Falque, V. M.; Fusar, F.; Gardiner, M.; Hamelin, E. M. N.; Hesterkamp, T.; Ichihara, O.; Jones, R. S.; Mather, O.; Mercurio, C.; Minucci, S.; Montalbetti, C. A. G. N.; Muller, A.; Patel, D.; Phillips, B. G.; Varasi, M.; Whittaker, M.; Winkler, D.; Yarnold, C. J. *ChemMedChem* **2009**, *4*, 963.
41. Feldman, R. I.; Mintzer, B.; Zhu, D. G.; Wu, J. M.; Biroc, S. L.; Yuan, S. D.; Emayan, K.; Chang, Z.; Chen, D.; Arnaiz, D. O.; Bryant, J.; Ge, X. S.; Whitlow, M.; Adler, M.; Polokoff, M. A.; Li, W. W.; Ferrer, M.; Sato, T.; Gu, J. M.; Shen, J.; Tseng, J. L.; Dinter, H.; Buckman, B. *Chem. Biol. Drug Des.* **2009**, *74*, 43.
42. Richardson, C. J.; Barlow, D. J. *J. Pharm. Pharmacol.* **1996**, *48*, 581.
43. Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. *Nature* **1986**, *323*, 533.
44. Ertl, P.; Rohde, B.; Selzer, P. *J. Med. Chem.* **2000**, *43*, 3714.
45. Janin, Y. L.; Pase, A. T. *Drug Discovery Today* **2010**, *15*, 342.
46. Wegele, H.; Muller, L.; Buchner, J. *Rev. Physiol., Biochem. Pharmacol.* **2004**, *151*, 1.