Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Molecular Graphics and Modelling 24 (2006) 328–340
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Classification and comparison of ligand-binding sites derived from grid-mapped knowledge-based potentials

Christian Hoppe [a,b], Christoph Steinbeck [b], Gerd Wohlfahrt [a,*]

[a] Orion Pharma, Medicinal Chemistry, P.O. Box 65, FIN-02101 Espoo, Finland
[b] University of Cologne, Cologne University Bioinformatics Center (CUBIC), Zülpicher Str. 47, D-50674 Köln, Germany

Received 4 May 2005; received in revised form 29 August 2005; accepted 29 September 2005
Available online 2 November 2005

Abstract

We describe the application of knowledge-based potentials implemented in the MOE program to compare the ligand-binding sites of several proteins. The binding probabilities for a polar and a hydrophobic probe are calculated on a grid to allow easy comparison of binding sites of superimposed related proteins. The method is fast and simple enough to simultaneously use structural information of multiple proteins of a target family. The method can be used to rapidly cluster proteins into subfamilies according to the similarity of hydrophobic and polar fields of their ligand-binding sites. Regions of the binding site which are common within a protein family can be identified and analysed for the design of family-targeted libraries or those which differ for improvement of ligand selectivity.

The field-based hierarchical clustering is demonstrated for three protein families: the ligand-binding domains of nuclear receptors, the ATP-binding sites of protein kinases and the substrate binding sites of proteases. More detailed comparisons are presented for serine proteases of the chymotrypsin family, for the peroxisome proliferator-activated receptor subfamily of nuclear receptors and for progesterone and androgen receptor. The results are in good accordance with structure-based analysis and highlight important differences of the binding sites, which have been also described in the literature.

Keywords: Molecular fields; Grid-based comparison; Knowledge-based potentials; Nuclear receptors; Protein kinases; Serine proteases

1. Introduction

Selectivity towards a biological target is an important property for a drug candidate in order to minimize potential side effects. Traditionally, this has been achieved by cycles of modification and testing of lead compounds. In the absence of structural information of the protein targets, ligand-based QSAR methods have been used to improve specificity, of which comparative molecular field analysis (CoMFA) [1,2] is one of the most successful. A drawback of this approach is that it requires a set of known active molecules with different specificities and whose three-dimensional structures have to be aligned in a meaningful way.

With the rapidly increasing number of protein structures, knowledge of the three-dimensional arrangement of ligand-binding sites became a valuable tool to guide drug design and to introduce receptor specificity early in the discovery process. Molecular fields derived from protein structures have been used to classify and to compare the binding sites of different related receptors [3–5]. These fields were calculated, e.g. with the GRID program [6] using probes whose interaction energies are defined by empirical force fields. Non-grid-based mapping of protein sites has been performed, e.g. by MCSS [7,8], which optimizes the position and orientation of multiple probes in the binding sites. The computationally more demanding MCSS method gives more details than GRID as additional orientational information is provided [8], but as the probe positions are not fixed here, comparison with related receptors is more complex.

Besides empirical force fields, knowledge-based potentials have been proven to characterise receptor–ligand interactions in an appropriate way [9,10]. The use of empirical packing preferences and knowledge-based potentials to assess preferred binding sites in proteins is a well established concept; some

Abbreviations: AR, androgen receptor; LBD, ligand-binding domain; NR, nuclear receptor; PCA, principal component analysis; PLS, partial least squares; PPAR, peroxisome proliferator-activated receptor; PR, progesterone receptor; RMSD, root mean square deviation

* Corresponding author. Tel.: +358 10 4294786; fax: +358 10 4294682.
E-mail address: gerd.wohlfahrt@orionpharma.com (G. Wohlfahrt).

1093-3263/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2005.09.013

examples of this approach include the work of Thornton and co-workers [11], as well as the IsoStar [12] and SuperStar methods produced by Verdonk and co-workers [13]. The advantage of knowledge-based potentials is their ability to describe complex interactions influenced by entropic effects or many-body interactions, which are difficult to quantify with empirical force fields [14].

After mapping of the binding sites, different methods for comparison can be applied, which usually rely on superposition of related protein structures. Principal component analysis [4] or trend vector methods [15] can be applied to extract relevant differences between the fields of the receptors. The first method identifies the most variable features among all receptors in a reduced descriptor space, while the second one finds contour levels above chance correlations from a vector in the original descriptor space.

Most studies have been focussed on the identification of regions which differ between receptors in order to improve selectivity of a ligand, but regions which are common within a protein family are also of interest for the design of family-targeted libraries or to support identification of privileged substructures.

In the present paper, we describe the application of knowledge-based potentials implemented in MOE [16], which use experimental contact statistics fitted to analytical functional forms to identify specific interactions with a protein structure. These potentials include besides distance-dependent also angle and out-off plane dependent distributions. MOE contact statistics have already successfully been used to help refine results from molecular docking runs [17] on the NSAID/COX-2 system, to aid in biodistribution prediction [18] and to explain inhibitor–protein contacts in insect cytochrome P450 binding sites [19]. We calculated the binding probabilities for a polar and a hydrophobic probe on the intersection points of a grid to allow easy comparison of binding sites of superimposed related proteins. The method is fast and simple enough to simultaneously use structural information of multiple proteins of a target family. Using several structures of the same receptor helps to identify the most important interacting regions, which are, e.g. seen with all ligands. Compared to fields derived from a single protein structure this also reduces spurious results, which could be related to experimental inaccuracy or to flexible side-chains resulting in small differences among crystal structures of the same receptor. The method can be used to rapidly cluster proteins into subfamilies according to the similarity of hydrophobic and polar fields of their ligand-binding sites. Regions of the binding site which are common or differ within a protein family can be identified and analysed. Knowledge about common regions is, e.g. useful for the design of family-targeted libraries and differences can be used to improve selectivity of a ligand.

The field-based clustering method is demonstrated for three protein families containing many pharmaceutically relevant targets: the LBDs of nuclear receptors, the ATP-binding sites of protein kinases and the substrate binding sites of proteases [4,5,20,21].

Three serine proteases from the chymotrypsin family are used in the test set. Two of them, thrombin and factor Xa, are involved in the blood clotting cascade and are therefore important targets in the development of anticoagulant or antithrombotic drugs. Trypsin is a pancreatic enzyme involved in digestion. Therefore, selectivity for thrombin and factor Xa over trypsin would improve bioavailability and minimize side effects [4,22].

Most proteins of the nuclear receptor superfamily (NR) act as ligand-activated transcription factors, but the exact mechanism by which the nuclear receptors affect gene transcription is still poorly understood, as is in many cases the role of the subfamilies and their subtypes [2]. Despite the low sequence identity between the LBDs of different NR subfamilies, all NRs share a similar fold and many can bind a range of similar ligands. Depending if the bound ligand is agonistic or antagonistic, the carboxyl-terminal helix H12 is found in either one or another orientation. In the agonist-bound conformation H12 closes the ligand-binding site and shields it from the solvent, whereas in the antagonist-bound conformation H12 does not close the binding pocket. This leads to rather large differences between the properties of the binding sites in the two conformations. A detailed pairwise comparison is presented for the progesterone and androgen receptor, whose binding sites are very similar. High selectivity for only one of the closely related androgen, progesterone, glucocorticoid or mineralocorticoid receptors is important in order to reduce side effects of drug candidates [23,24].

The superfamily of eukaryotic protein kinases is formed of homologous proteins related by their catalytic domains. Although they may have different regulation modes or substrate specificities, they share a common catalytic core structure, which indicates how phosphate is transferred from the kinase to a hydroxyl group in the protein substrate [5]. Kinases play an important role in diverse biological processes such as controlling, signalling and triggering a broad variety of cellular events. Of pharmaceutical interest is the possibility of inhibiting the ATP binding site [5,25–29]. A problem with this approach is that, besides the different kinase subfamilies, more than 2000 ATP-utilizing proteins are estimated in the human genome.

These examples of NRs, kinases and proteases illustrate that methods for analyzing subfamilies or improving subtype specificity of ligands are important in the development of compounds with fewer side effects.

2. Methods

2.1. Protein structures

The protease dataset consists of 13 protein X-ray structures and was taken from literature [30] (Table 1). Sixty-seven nuclear receptor X-ray structures from three subfamilies [31] were taken from the NucleaRDB [32] (Table 2). The kinase dataset was retrieved from the PDB [33] using the search criteria human, X-ray, resolution equal or lower than 2.5 Å and the datasets from Deng et al. [34] and Naumann and Matter [5] (Table 3). Overall 75 protein kinase structures were chosen.

Table 1
List of crystal structures used in the serine protease test set

Family/PDB code	Organism[a]	*R* (Å)	Ligand name	Ca	Ca (%)	RMSD Ca
Alpha-thrombin						
1dwd[b]	HU	3.0	NAPAP	REF	REF	REF
1dwb	HU	3.2	Benzamidine	257	100	0.000
1dwc	HU	3.0	MD-805	257	100	0.121
1qur	HU	2.0	Benzamidine	257	100	0.330
7kme	HU	2.1	SEL2711	248	96.5	0.319
Coagulation factor Xa						
1fax	HU	3.0	–	203	79.0	0.816
1hcg	HU	2.2	–	204	79.4	0.820
1xkb	HU	2.4	FX-2212A	205	79.8	0.838
Trypsin						
1ppc	BO	1.8	NAPAP	199	77.4	0.752
1mts	BO	1.9	–	200	77.8	0.743
1mtu	BO	1.9	–	200	77.8	0.743
1mtv	BO	1.9	–	200	77.8	0.758
1mtw	BO	1.9	–	198	77.0	0.729

R is the resolution of the X-ray structure. Ca refers to the number and percentage of Cα-backbone atoms, which are aligned by the program BRAGI to the given RMSD.
[a] HU, human; BO, bovine.
[b] Reference structure for superposition.

Table 2
List of 67 crystal structures of nuclear receptors used in the analysis

Family/PDB code	Organism[a]	*R* (Å)	Inhibition mechanism[b]	Ligand name	Ca	Ca (%)	RMSD Ca
RXR-alpha							
1dkf	HU/MO	2.5	ANT	OLI	142	57.9	1.016
1fby	HU	2.2	AG	REA	147	60.0	1.103
1fm6	HU	2.1	AG	REA	149	60.8	1.029
1fm9	HU	2.1	AG	REA	124	50.6	0.971
1k74	HU	2.3	AG	REA	147	60.0	1.036
1mv9	HU	1.9	AG	HXA	151	62.0	0.965
1mvc	HU	1.9	AG	BMS649	152	62.1	0.990
1mzn	HU	1.9	AG	BMS649	165	63.3	1.074
RAR-y							
1exa	HU	1.6	AG	BMS270394	135	55.1	1.099
1exx	HU	1.6	AG	BMS270395	137	55.9	1.139
1fcx	HU	1.5	AG	BMS184394	137	55.9	1.114
1fcy	HU	1.3	AG	CD564	138	56.3	1.120
1fcz	HU	1.3	PAG	BMS181156	135	55.1	1.103
1fd0	HU	1.3		SR11254	134	54.6	1.091
PR							
1a28	HU	1.8	AG	STR	164	66.9	0.963
1e3k	HU	2.8	AG	Metribolone	174	71.0	1.022
ER-a							
1a52	HU	2.8	AG	EST	219	90.0	0.441
1ere	HU	3.1	AG	EST	232	94.5	0.438
1err	HU	2.6	ANT	RAL	211	86.1	0.487
1g50	HU	2.9	AG	EST	235	95.9	0.473
1gwq	HU	2.4	AG	ZTW	236	96.3	0.448
1gwr	HU	2.4	AG	EST	232	94.6	0.419
1l2i	HU	1.9	AG	ETC	233	95.1	0.291
1pcg	HU	2.7	AG	EST	222	91.3	0.371
1qkt	HU	2.2	AG	EST	209	85.3	0.430
1qku	HU	3.2	AG	EST	235	95.9	0.463
1uom	HU	2.8	ANT	PTI	205	84.6	0.407
3erd[c]	HU	2.0	AG	DES	REF	REF	REF
3ert	HU	1.9	ANT	OHT	204	83.2	0.505
ER-b							
1l2j	HU	2.9	ANT	ETC	184	75.1	0.910

Table 2 (*Continued*)

Family/PDB code	Organism[a]	*R* (Å)	Inhibition mechanism[b]	Ligand name	Ca	Ca (%)	RMSD Ca
1nde	HU	3.0	ANT	MON	191	77.9	0.842
1qkm	HU	1.8	ANT	GEN	194	78.1	0.827
1qkn	RA	2.2	ANT	RAL	201	82.0	0.718
AR							
1e3g	HU	2.4	AG	Metribolone	171	69.8	0.963
1gs4	HU	1.9		ZK5	169	60.9	0.901
1i37	RA	2.0	AG	DHT	171	69.8	0.939
1i38	RA	2.0	AG	DHT	171	69.8	0.948
PPAR-a							
1i7g	HU	2.2	AG	AZ242	113	46.1	1.103
1k7l	HU	2.5	AG	GW409544	126	51.4	1.203
1kkq	HU	3.0	ANT	GW6471	116	47.7	1.157
PPAR-y							
1fm6	HU	2.1	AG	BRL	129	56.6	1.164
1fm9	HU	2.1	AG	GI262570	132	53.9	1.135
1i7i	HU	2.3	AG	AZ242	120	48.9	1.159
1k74	HU	2.3	AG	GW409544	128	52.2	1.135
1knu	HU	2.5	AG	YPA	127	51.8	1.182
2prg	HU	2.3	AG	BRL	132	53.8	1.101
4prg	HU	2.9	PAG	0072	131	53.4	1.196
PPAR-d							
1gwx	HU	2.5	AG	GW2433	108	44.1	1.206
2gwx[d]	HU	2.3	–	–	–	–	–
3gwx	HU	2.4	AG	EPA	107	43.6	1.101
PXR							
1ilh	HU	2.7	AG	SRL	133	54.2	1.149
VDR							
1db1	HU	1.8	AG	VDX	149	60.8	1.088
1ie8	HU	1.5	AG	KH1	149	61.0	1.110
1ie9	HU	1.4	AG	VDX	147	60.0	1.089
TR-b							
1bsx	HU	3.7	AG	T3	132	53.8	1.101
1nax	HU	2.7	AG	IH5	122	50.2	1.154
1nq1	HU	2.9	AG	4HY	139	57.0	1.149
1nq2	HU	2.4	AG	4HY	132	54.3	1.137
1nuo	HU	3.1	AG	4HY	119	48.9	1.134
1n46	HU	2.2	AG	PFA	131	53.5	1.141
ERR3							
kv6	HU	2.7	AG	DES	173	70.1	0.805
ROR-a							
1n83	HU	1.6	AG	Cholesterol	131	53.4	1.176
ROR-b							
1k4w	RA	1.9	AG	STE	140	57.1	1.181
HF-4y							
1lv2	HU	2.7		PLM	146	59.5	1.037
GCR							
1m2z	HU	2.5	AG	DEX	163	66.5	0.979
1nhz	HU	2.3	ANT	486	160	63.3	1.046
LXR-b							
1p8d	HU	2.8	AG	CO1	150	62.5	1.146
PregnaneX							
1ilh	HU	2.7	AG	SR12813	135	51.8	1.161

R is the resolution of the X-ray structure. Ca refers to the number and percentage of Cα-backbone atoms, which are aligned by the program BRAGI to the given RMSD.

[a] HU, human; BO, bovine; RA, rat; MO, mouse.

[b] ANT is an antagonistic and AG is an agonistic ligand or receptor conformation.

[c] Reference structure for superposition.

[d] Used in the PPAR study.

Table 3
List of 75 crystal structures of protein kinases used in the analysis

Family/PDB code	Organism[a]	*R* (Å)	Ligand name	Ca	Ca (%)	RMSD Ca
PKA						
1atp[b]	MO	2.2	ATP	REF	REF	REF
1bx6	MO	2.1	Balanol	311	92.6	0.779
1cdk	RAB	2.0	AMPPNP	336	100	0.362
1fmo	MO	2.2	ADP	336	100	0.356
1jbp	MO	2.2	ADP	336	100	0.361
1l3r	MO	2.0	ADP	336	100	0.363
1ydr	BO	2.2	H7+	332	98.8	0.360
1yds	BO	2.2	H8+	332	98.8	0.329
1ydt	BO	2.3	H89+	332	98.8	0.349
CDK2						
1aq1	RA	2.5	Staurosporine	174	51.7	1.059
1b38	HU	2.0	ATP	170	50.6	1.127
1b39	HU	2.1	ATP	171	50.8	1.139
1ckp	HU	2.0	Purvalanol	170	50.6	1.102
1di8	HU	2.2	–	171	50.8	1.141
1dm2	HU	2.1	Hymenialdisine	176	52.4	1.062
1e1v	HU	1.9	NU2058	153	45.5	1.170
1e1x	HU	1.8	NU6027	167	49.7	1.179
1e9h	HU	2.5	Indirubin	160	47.6	1.106
1fin	HU	2.3	ATP	157	46.7	1.113
1fvt	HU	2.2	–	179	53.3	1.146
1gii	HU	2.0	1PU	173	51.5	1.172
1gij	HU	2.2	2PU	168	50.0	1.180
1gz8	HU	1.3	MBP	105	56.7	1.168
1hck	HU	1.9	ATP	168	50.0	1.128
1jsv	HU	2.0	U55	170	50.6	1.139
1ke5	HU	2.2	LS1	172	51.2	1.172
1ke6	HU	2.0	LS2	172	51.2	1.157
1ke7	HU	2.0	LS3	174	51.8	1.177
1ke8	HU	2.0	LS4	171	50.8	1.160
1ke9	HU	2.0	LS5	172	51.2	1.172
1oir	HU	1.9	HDY	171	50.8	1.105
1oit	HU	1.6	HDT	171	50.8	1.166
1qmz	HU	2.2	ATP	170	50.6	1.110
CaMK-I/CDK1						
1a06	RA	2.5	–	149	53.4	1.120
CK-1 alpha						
1csn	BA	2.0	ATP	173	51.5	1.147
2csn	BA	2.5	CKI7	172	51.2	1.210
CK-2 alpha						
1daw	MA	2.2	AMPPNP	173	51.5	1.073
1day	MA	2.2	GMPPNP	171	50.9	1.092
1j91	MA	2.2	TBS	172	51.2	1.087
1lp4	MA	1.8	AMPPNP	174	51.8	1.085
MAP/ERK2						
1erk	RA	2.3	–	162	48.2	1.105
1gol	RA	2.8	ATP	161	47.9	1.103
3erk	RA	2.1	SB220025	162	47.0	1.121
P38						
1a9u	HU	2.5	SB203580	148	44.05	1.113
1bl6	HU	2.5	SB216995	141	41.9	1.102
1bl7	HU	2.5	SB220025	141	41.9	1.090
1bmk	HU	2.4	SB218655	141	41.9	1.090
1kv1	HU	2.5	BMU	136	40.5	1.106
1p38	MO	2.1	–	144	42.9	1.096
1pme	HU	2.0	SB202190	158	47.0	1.119
Phosphorylase kinase						
1phk	RAB	2.2	ATP	216	64.3	0.900
1ql6	RAB	2.4	ATP	217	64.6	0.889

Table 3 (*Continued*)

Family/PDB code	Organism[a]	*R* (Å)	Ligand name	Ca	Ca (%)	RMSD Ca
2phk	RAB	2.6	ATP	221	65.8	1.166
DAP						
1ig1	HU	1.8	AMPPNP	210	62.5	1.001
1jkk	HU	2.4	AMPPNP	210	62.5	1.008
1jkl	HU	1.6	AMPPNP	212	63.1	1.006
IRK						
1gag	HU	2.7	112	137	40.7	1.122
1i44	HU	2.4	AMPPNP	114	33.9	1.215
1ir3	HU	1.9	AMPPNP	136	40.5	1.137
FGFIR						
1agw	HU	2.4	SU4984	122	36.3	1.130
1fgi chainA	HU	2.5	SU5402	120	35.7	1.217
1fgi chainB	HU	2.5	SU1	123	36.6	1.225
c-Src						
1ksw	HU	2.8	AMPPNP	140	41.7	1.212
2src	HU	1.5	AMPPNP	146	43.5	1.219
1GF-IR						
1jqh	HU	2.1	PP1	141	41.9	1.193
1k3a	HU	2.1	ACP	139	41.3	1.129
HCK						
1ad5	HU	2.6	AMPPNP	150	44.6	1.184
1qcf	HU	2.0	PP1	141	41.9	1.193
LCK						
1qpc	HU	1.6	AMPPNP	147	43.8	1.161
1qpd	HU	2.0	Staurosporine	149	44.4	1.141
3lck	HU	1.7	–	132	39.3	1.148
1qpe	HU	2.0	PP2	141	41.9	1.146
1qpj	HU	2.2	Staurosporine	159	47.3	1.124
ABL						
1fpu	MO	2.4	STI-571	141	41.9	1.136
1iep	MO	2.1	STI-571	140	41.7	1.155

R is the resolution of the X-ray structure. Ca refers to the number and percentage of Cα-backbone atoms, which are aligned by the program BRAGI to the given RMSD.

[a] HU, human; BO, bovine; RA, rat; MO, mouse; RAB, rabbit; BA, bacteria.

[b] Reference structure for superposition.

Cofactors, counter-ions, structural water and additional peptide chains were removed. Water was excluded in the comparison as not all structures contain all equivalent water molecules. This would be a source of additional noise in the analysis.

2.2. Superposition of protein structures

All protein structures were superimposed on a reference protein structure. As reference protein for the protease dataset, we used alpha-thrombin (1dwd), estrogen receptor alpha in an agonistic conformation with diethylstilbestrol as ligand (3erd) for the NR dataset and for the kinase dataset a cyclic adenosine 5′-monophosphate (cAMP)-dependent protein kinase (1atp). For the superposition of alpha carbons, the method of Lessel and Schomburg [35] was used with default parameters as implemented in the program BRAGI [36]. From structures with two chains of identical amino acid sequence, only the first one was considered. If there were two different chains, as in heterodimers, both chains were used.

2.3. Knowledge-based interaction fields of ligand-binding sites

All calculations were performed with the MOE program package [16]. The interaction field calculation and the binding site comparison were done with a script in SVL (the native Scientific Vector Language of MOE). First, a consensus binding site for the aligned structures was defined in such a way that each protein atom within 4.5 Å from the corresponding ligands belongs to the binding site. A grid with a lattice constant of 0.5 Å was spanned over this consensus binding site. Different grid-spacing has been tested. Spacing of 1 Å led to a rather broad distribution of properties resulting in not readily interpretable pictures. Grid-spacing of 0.25 Å was impractical for technical reasons as the field maps became very large. The clustering process took about 10-fold longer and produced the same results as with 0.5 Å spacing.

In the next step, a knowledge-based contact potential for two probes was calculated on each grid point for all aligned protein structures (ligands and crystal water were excluded). A

hydrophobic and a polar probe were used as implemented in the contact statistics method of MOE. The polar probe is derived from nitrogen and oxygen atoms capable of forming hydrogen bonds or metal interactions, whereas the hydrophobic type consists of all non-polar atoms like carbon, sulphur and halogens. The resulting contact probabilities for the two probes on the grid points are saved into vectors, with the position in each vector corresponding to the same grid point in different proteins.

The interaction fields can be calculated in two ways: only one structure is used for calculating the contact probability fields, or the contact probability field is calculated from several aligned protein structures, of either the same or different receptors, which we will subsequently call the ''assembly-set approach''. Only one contact probability field is calculated for each probe. To analyse differences and similarities of the hydrophobic and polar fields between different subfamily structures, the corresponding fields are subtracted from each other for so-called difference fields or added for so-called similarity fields:

$$\text{differenceField}[i] = \text{fieldOne}[i] - \text{fieldTwo}[i]$$

$$\text{similarityField}[i] = \begin{cases} 1 - (\text{abs}(\text{fieldOne}[i] - \text{fieldTwo}[i])), \\ \quad \text{if fieldOne}[i] \geq 0.9 \,\&\, \text{fieldTwo}[i] \geq 0.9 \\ 0 \end{cases}$$

where i is the actual grid point and differenceField, similarityField, fieldOne, fieldTwo are vectors of double values. Only probabilities were accepted with an absolute value equal or greater than 0.9 or 1.8 for both field types, respectively. In an initial analysis, we found that cut-off values of 0.85 produce too many difference fields in order to obtain easily interpretable pictures. Using cut-offs of 0.95 produced very few similar regions even for the comparison of closely related receptors. Clustering the protein families using these values did not result in different subfamilies than with a cut-off of 0.9.

The geometric centres of the resulting probability fields can be calculated for the construction of receptor-based pharmacophores or for further visual analysis.

For comparisons of receptor families, the interaction fields of the two probes were calculated for several aligned protein structures (assembly-set approach). Only probabilities were accepted with a value equal or greater than 0.9 and with at least four other significant values within 1 Å radius around the respective grid point. Otherwise, the probability for a grid point was set to zero in order to improve the signal–noise ratio by suppressing areas of weak or unspecific interactions. The grid-based probability fields were inspected visually and displayed with MOE.

2.4. Clustering of the interaction fields

The interaction fields of hydrophobic or polar probes for single structures can be clustered to show the relationships among the different subtypes and receptor families. A consensus grid was calculated from the hydrophobic and polar probe grids. For this purpose, only probabilities equal or higher than 0.9 were taken into account, all other probabilities were set to zero. Hydrophobic probe probabilities were set positive, polar probe probabilities negative. No scaling was applied to any of the two fields. Clustering of these consensus grids was performed with an integrated pair of programs Cluster 3.0/TreeView [37]. Distance measure was based on the Pearson correlation and a hierarchical complete linkage clustering was then produced.

3. Results

Before applying our method to clustering of larger protein families and detailed pairwise comparison of selected members, we tried to estimate its performance for two field-based comparisons of binding sites described in the literature: serine proteases from the chymotrypsin family [4,30] and the PPAR subfamily of nuclear receptors [21].

3.1. Serine proteases

In this example, we analyse the differences between thrombin and two other serine proteases, trypsin and factor Xa (Table 1). Fig. 1 shows the hierarchical single linkage clustering of the polar and hydrophobic fields of their binding sites. Field-based clustering groups the proteins in the same way as sequence-based clustering (not shown).

Fig. 2 shows the differences between the hydrophobic and polar probe fields in the binding sites of trypsin and factor Xa in respect to thrombin. The differences in the hydrophobic fields of factor Xa and trypsin (green and red mesh) near residue 99 in the D pocket can be explained by a loop in thrombin containing residues 96–98, which reduces the space in this region of the binding site (Fig. 2). Furthermore, in factor Xa a Phe is in position 174, but in trypsin there is a Gln, which is indicated by a large hydrophilic difference field (red solid) in this area of the factor Xa binding site.

In the P pocket, Tyr60a in thrombin has no counterpart in the other two enzymes—reflected by the corresponding hydrophilic difference fields (green solid). This is similar for Tyr99 in factor Xa. This mutation blocks a part of the binding site, which is accessible in thrombin and trypsin. This is indicated by the corresponding hydrophobic difference fields (red and green mesh). Two additional differences seen in the hydrophobic fields in the S1 pocket correspond to the mutation of Glu192 in thrombin to Gln in trypsin and factor Xa, respectively, to the mutation of Ser190 in trypsin to Ala190 in thrombin and factor Xa.

Differences and similarities displayed are in good accordance with the literature [4,15] and are able to explain different specificities of certain inhibitors.

3.2. PPAR family

The second system we analyse is the PPAR family (Table 2). Fig. 3 shows the hierarchical clustering of the molecular fields

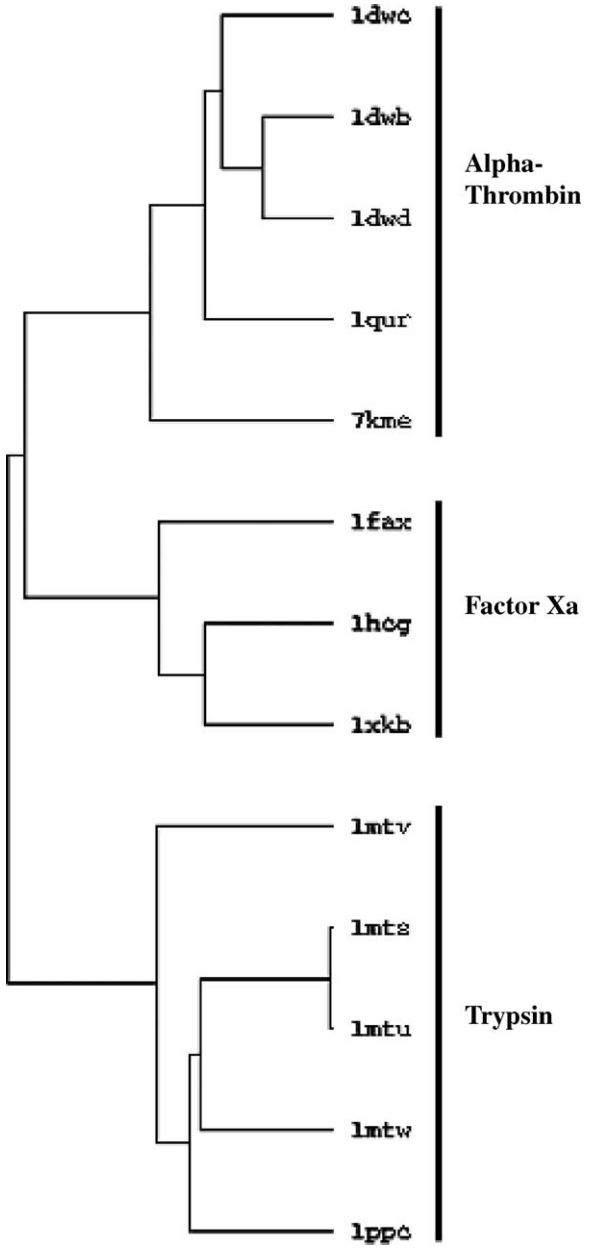

Fig. 1. Hierarchical complete single linkage clustering of 13 serine proteases according to the molecular fields derived from their ligand-binding sites. The Pearson correlation coefficient was used as distance measure.

of the binding sites of PPARα, PPARγ and PPARδ in complex with different ligands. The clustering groups the PPARα structure 1i7g in the PPARδ cluster and the PPARγ structure 4prg in the PPARα cluster. 1i7g is an agonistic conformation whereas the PPARα structure 1kkq is an antagonistic one. 2gwx has no ligand bound, so that the binding site is more similar (closed site) to the 1i7g structure than to the antagonistic 1kkq one (open site). 4prg is a structure with a partial agonist and the receptor conformation is closed like with full agonists, but the ligand occupies a similar volume to the antagonistic ligand in the 1kkq structure. These findings are in accordance with the literature [21].

In Fig. 4, the difference fields derived from the PPAR-binding sites indicate the importance of the following regions.

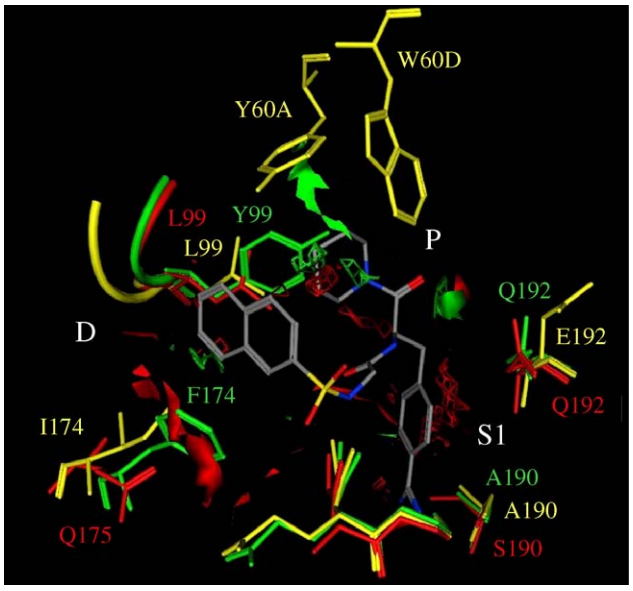

Fig. 2. The hydrophobic (mesh) and polar (solid) difference fields between thrombin and trypsin (red) and thrombin and factor Xa (green). The amino acids in the binding-site are coloured red (trypsin), green (factor Xa) and yellow (thrombin). The inhibitor NAPAP is shown in grey. The pockets are labelled in white.

The reduction of the size of the distal pockets because of mutations in positions 264 and 284 for PPARδ (left distal) and for PPARα and γ of residues 281, 348, 339 (left distal) and 288 (right distal) is reflected by the location of the hydrophobic and hydrophilic difference fields (see Fig. 4, mesh fields below Val339). Also an exchange in position 364 is mentioned in the literature [21], which affects the shape of the left distal pocket, but this is not obvious from the fields shown in Fig. 4.

The effect of the exchange in position 453 (Leu to Met) in PPARδ, which causes steric hindrance at the entrance of the binding pocket, is seen in the hydrophilic difference field in the head region (red solid). This hydrophilic difference field is also influenced by a mutation of His323 (PPARδ) to Tyr (PPARα). Changes in the hydrophobic and hydrophilic difference fields (red mesh and solid) for the mutations of residues 292 and 323 can be seen in the right distal pocket and linker region, respectively. Furthermore, we found a hydrophilic difference field near Gln259 that originates from differences in its side-chain orientation. In addition to this, we find differences in the hydrophobic fields near residues 268 and 270 (yellow mesh), caused by amino acids located in the second sphere of the protein up to 5 Å away from the ligand. This difference is not mentioned in the literature, but in general the results are in good accordance with other publications [21].

3.3. Clustering of protein families

Figs. 5 and 6 show the results of hierarchical clustering using the hydrophobic and polar consensus fields of 75 protein kinases and 67 NRs, respectively.

In the kinase dataset (Table 3), the cdk2 family was grouped into four clusters (Fig. 5). The first cluster consists of 13 cdk2 structures without the cyclin A ligand so that the activation loop

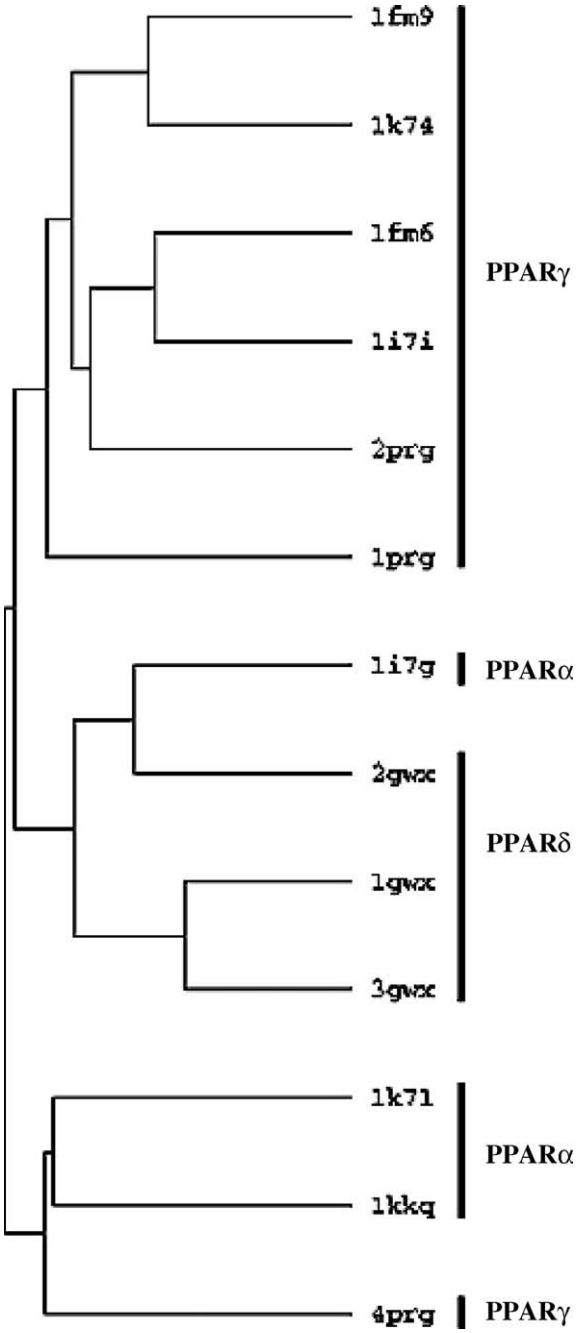

Fig. 3. Hierarchical complete single linkage clustering of 13 PPARs according to the molecular fields derived from their ligand-binding sites. The Pearson correlation coefficient was used as distance measure.

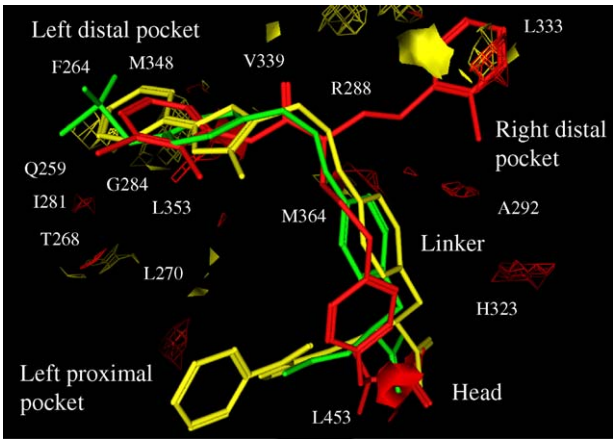

Fig. 4. The hydrophobic difference fields (mesh) and polar difference fields (solid) of PPARγ (yellow) and PPARδ (red) in respect to PPARα. The inhibitors from the X-ray structures 1fm9 (PPARγ), 1i7g (PPARα) and 1gwx (PPARδ) are shown in yellow, green and red, respectively. For clarity, only approximate positions of amino acid side-chains are indicated by the residue numbers from the PPARγ structure (1fm9).

is oriented towards the binding site. This is the same for the structures in the last cluster with eight members (Fig. 5), but these two clusters have different binding modes of their ligands. The ligands of the last cluster are located in the so-called hydrophobic pocket of the ATP binding site [38]. Each of the other two clusters consists of only one member (1e9h and 1qmz). Neither is complexed with cyclin A and the two structures differ in their binding modes. In 1qmz, ATP binds to the hydrophobic pocket whereas in 1e9h it binds as in the first cdk2 family cluster. This illustrates that not only changes in the receptor conformation due to the orientation of the activation loop can be differentiated by our method, but also those caused by different binding modes of ligands. This information could be used to classify ligands according to the changes they induce to the receptor, which might be linked to functional aspects.

In the NR family (Table 2), no clear separation could be found between the ligand-binding sites of estrogen receptors α and β (Fig. 6). Their binding sites differ by only two conservative exchanges of aliphatic side-chains [39]. Despite a hydrophobic residue blocking, a large part of the binding pocket of the orphan receptor ERR, which should result in rather different molecular fields, its binding site was found to be most related to that of the two estrogen receptors. This corresponds well to its sequence-based classification as estrogen related receptor and its ability to bind a range of estrogenic compounds [40].

The information obtained with the hydrophobic and polar probe seems to be sufficient for a rapid discrimination as to which subfamily a given NR or kinase belongs, as the families are similar to sequence- or function-based classification of kinases [5] and NRs [31].

3.4. Pairwise comparison of androgen and progesterone receptors

In the following, we will discuss results of the assembly-set approach for the comparison of androgen and progesterone receptors. Residues, which are different in the binding sites of the two receptors, are highlighted in Fig. 7a. The fields for the hydrophobic probe (Fig. 7a) and for the polar probe (Fig. 7b) are well localized and mostly separated from each other, which helps to identify specific interactions, differences and similarities between the two receptors. The hydrophobic similarity field shows the general preference of the pockets for hydrophobic ligands (Fig. 7a, red), but also two conserved regions with a preference for polar groups are seen close to the C3 and C17 substituents of the steroidal ligands (Fig. 7b, red).

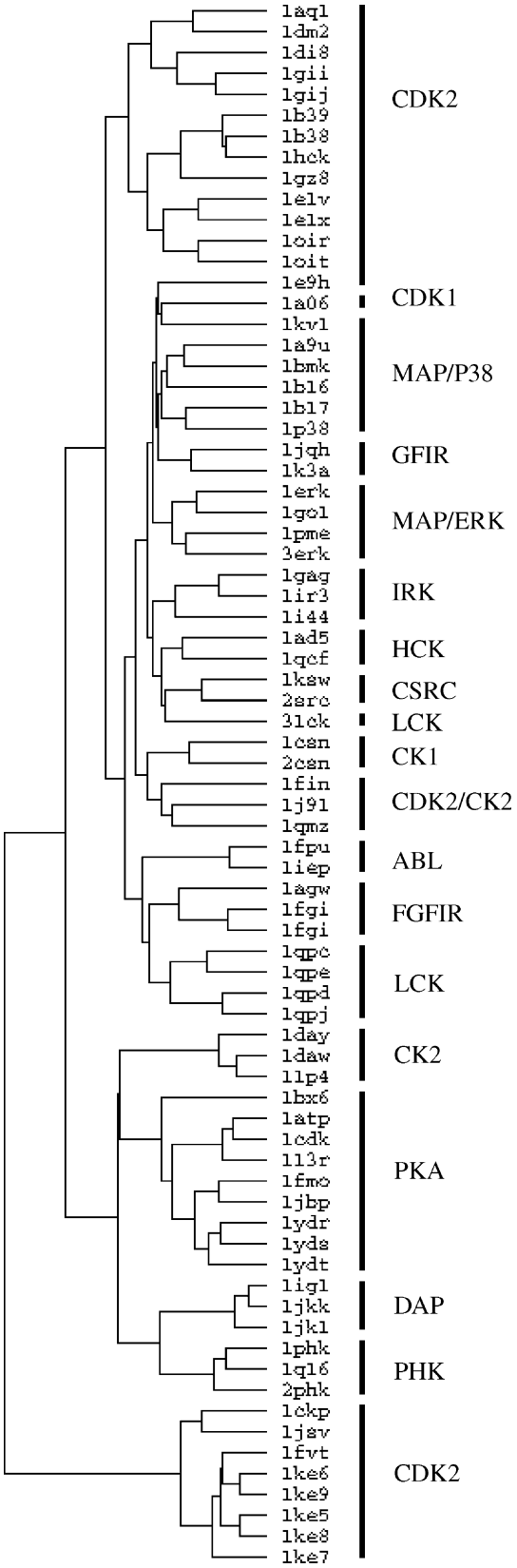

Fig. 5. Hierarchical complete single linkage clustering of 75 protein kinases according to the molecular fields derived from their ligand-binding sites. The Pearson correlation coefficient was used as distance measure.

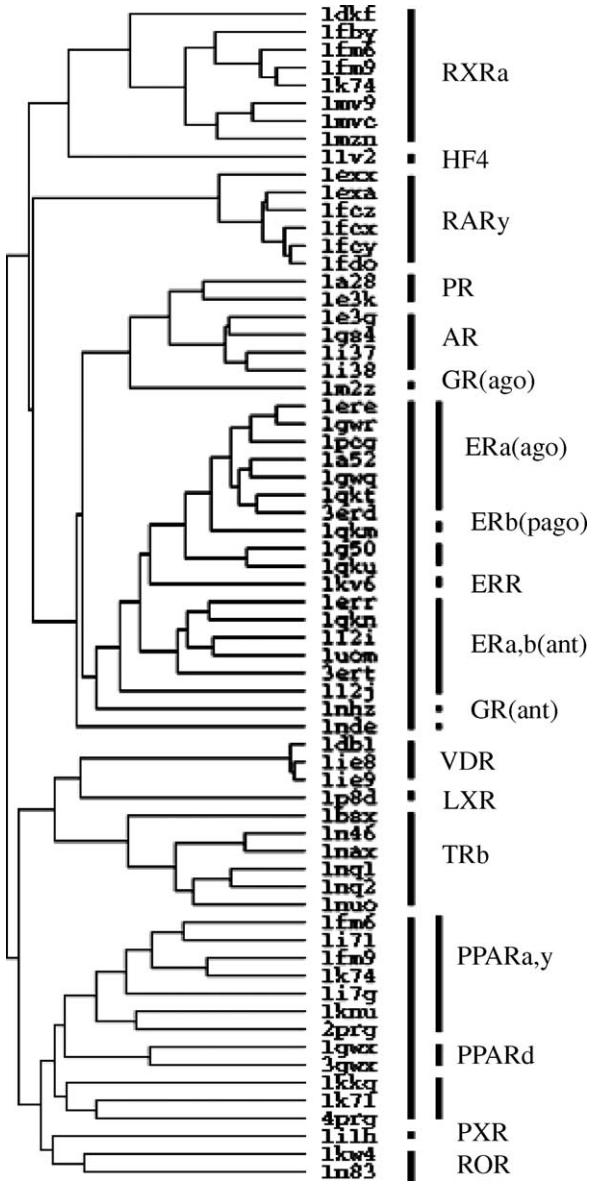

Fig. 6. Hierarchical complete single linkage clustering of 67 nuclear receptors according to the molecular fields derived from their ligand-binding sites. The Pearson correlation coefficient was used as distance measure.

This pattern corresponds well to the orientation of steroidal ligands in the binding pocket seen in the X-ray structures. The hydrophobic difference field shows the differences between progesterone and androgen receptor near residue Thr894(PR)/Leu880(AR). This difference is caused by higher polarity of Thr877 in AR compared to Cys891 in PR and the larger space occupied by the hydrophobic side-chain of Leu880 in AR, and has been used to explain the specificity for certain ligands [41].

The main difference in the polar fields of the two receptors is also found near Thr894(PR)/Leu880(AR). The polar probe contour for the androgen receptor is slightly smaller due to the hydrophobic side-chain of Leu880 and closer to Thr877. Both receptors show a positive polar contour at 2.1 Å distance from the backbone carbonyl group of Leu887(PR)/Leu873(AR) and 3.8 Å distance from the backbone nitrogen of Cys891, which is not targeted by the present ligands.

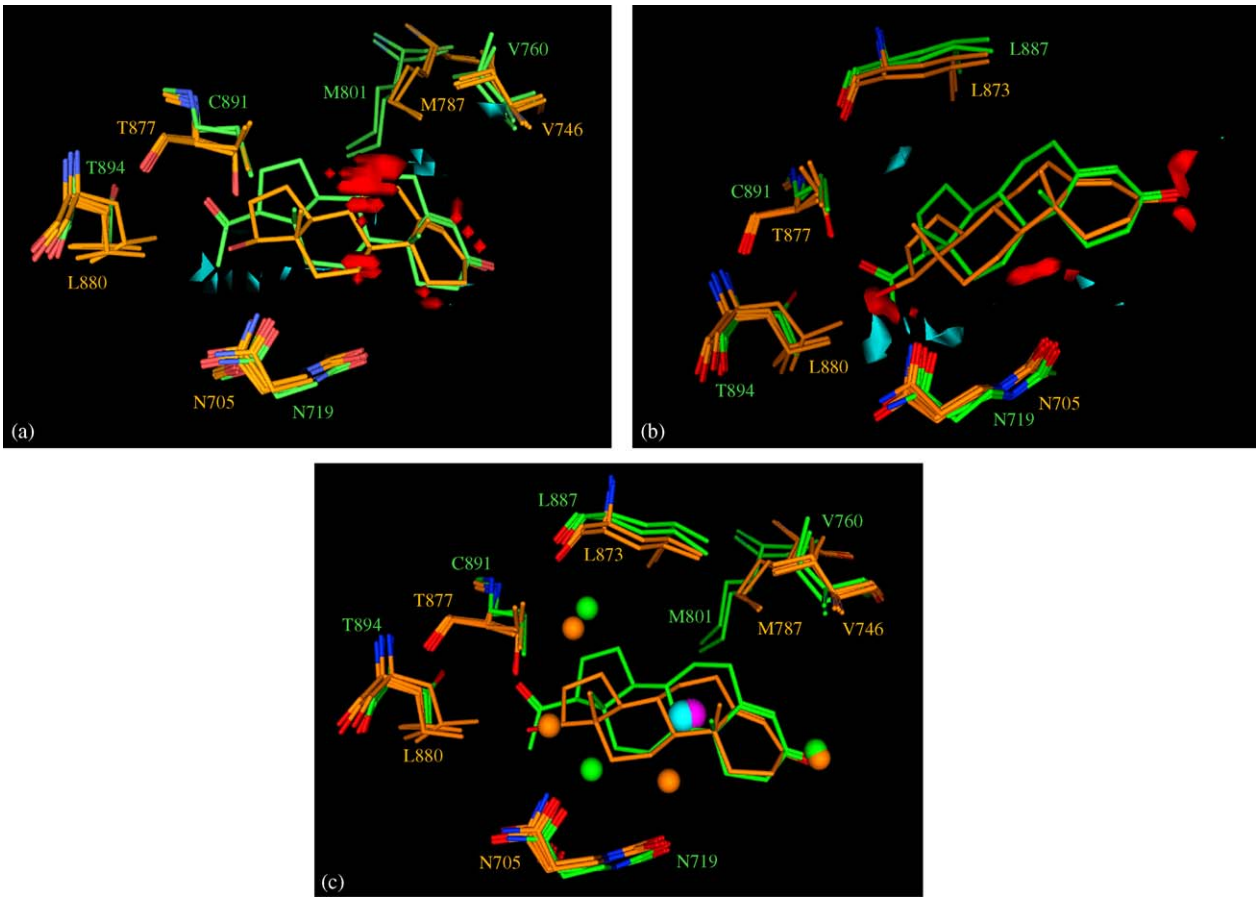

Fig. 7. Comparison of the ligand-binding sites of androgen and progesterone receptor. The residues are numbered according to the progesterone receptor (1a28) and the androgen receptor (1i37) structures. For clarity, only the ligands progesterone (green) and 5α-dihydro-testosterone (brown) are included. (a) The hydrophobic difference (cyan) and hydrophobic similarity (red) fields obtained by the assembly-set approach, showing the binding site and the ligands of the androgen receptor crystal structures (1e3g, 1gs4, 1i37, 1i38) in brown and the progesterone receptor crystal structures (1a28, 1e3k) in green. (b) The polar difference (cyan) and polar similarity (red) fields of AR and PR. (c) The geometric centres of the polar probe (PR green, AR brown) and hydrophobic probe (PR cyan, AR magenta) fields obtained by the assembly-set approach.

Important ligand–receptor interactions and differences in these regions are identified by our method [41,42]. The difference between the two NR binding sites originates mainly from the Cys891/Thr877 and Thr894/Leu880 mutations. This is also highlighted in a simpler way in Fig. 7c where the geometric centres of the polar and hydrophobic difference fields are displayed.

4. Discussion

We were able to show that mapping of binding sites with only two knowledge-based probes can produce a classification of the sites similar to more complex field-based methods [4,5,21,30]. The chosen combination of cut-off values seems already sufficient to cluster protein families and to detect ‘‘meaningful’’ differences without using statistical tools like PLS or PCA, which use scaling of different probes or produce large matrices. Noise is partly suppressed by using only points, which have high binding probabilities for a probe and which have also neighbours with high binding scores. Additionally, using several structures of the same receptor can further reduce spurious results originating from small differences between crystal structures, caused by experimental inaccuracy or different experimental conditions. It can also account for protein flexibility to some extent [43], if conformational changes do not distribute (smear) differences over a too large area, which would bring the signal under the cut-off value. One approach to solve this problem with multiple side-chain conformations would be the use of the same conformation for certain side-chains in all structures of one receptor [15]. In this way, important similarities between receptors will not get lost in the similarity maps. But on the other hand different side-chain conformations, induced by different ligands, point to areas, which can adjust to the ligands and this information could be useful for compound design.

The use of only two probes in original descriptor space creates maps which are readily and intuitively interpreted by visual inspection. For all examples described here, the differences and similarities seem to be consistent with direct comparison of the protein structures and with conclusions described in the literature.

Shifts in the superposition of the protein structures can lead to adjacent difference fields of opposite sign (preference). Those areas can be often distinguished from real signals by

visual inspection or should be interpreted at least with caution (see, e.g. Fig. 7b below the A ring of the steroids). As most methods which compare field maps, like CoMFA [1], depend very much on the superposition of the molecules; here the use of a field-based superposition algorithm [44] might bring some improvement. Another approach to overcome the alignment problem would be the use of alignment-independent descriptors [45]. Therefore, interaction fields derived from single protein structures could be converted into alignment-independent descriptors, e.g. by the method described by Pastor et al. [45]. Similarity of the binding sites would be measured based on these descriptors and applied to clustering. Back-projection of interesting features into the protein structure could be used for visualization and interpretation [46].

Improvement of the presented method could come from the implementation of a larger number of either knowledge-based [9,10] or GRID-like probes [6]. Differentiation between significant and non-significant fields might be sometimes problematic by using just the simple difference fields [4]. Here, the use of methods like CPCA or trend vector/PLS, which have been shown to identify relevant areas and differences in ligand-binding sites, could bring further improvement [4,15].

The focus of the present method lies on being simple, fast and robust in order to cluster and compare protein families with several members. Besides the identification of regions, which differ between receptors in order to improve selectivity of a ligand, regions which are common within a protein family are of interest for the design of family-targeted libraries or to support identification of privileged or consensus substructures of ligands. For certain indications, simultaneously targeting several related receptors can provide a superior therapeutic effect compared to the action of selective ligands [47].

For virtual screening of compound libraries pharmacophores can be generated from the probability fields. Therefore, geometric centres of the difference fields are calculated and the standard deviation is used as radius for the sphere. At the moment, no weights are assigned to the individual spheres, but they could be derived from different contact probabilities of the probes. This type of pharmacophore, derived from similarity or difference fields, could be also used as filters for docking procedures. Pharmacophoric points derived from similarity fields would express compound features, which are common within a receptor family, while points derived from difference fields focus the virtual screening process on more subtype-selective compounds.

Acknowledgement

We are grateful to Drs. Wolfram Altenhofen, Steve Maginn, Martti Ovaska and Chris Williams for helpful discussions and the revision of the manuscript.

References

[1] K.H. Kim, G. Greco, E. Novellion, A critical review of recent CoMFA applications, Perspect. Drug Discov. Des. 12 (1998) 257–315.
[2] P. Wolohan, D.E. Reichert, CoMFA and docking study of novel estrogen receptor subtype selective ligands, J. Comput. Aided. Mol. Des. 17 (2003) 313–328.
[3] M. Pastor, G. Cruciani, A novel strategy for improving ligand selectivity in receptor-based drug design, J. Med. Chem. 38 (1995) 4637–4647.
[4] M.A. Kastenholz, M. Pastor, G. Cruciani, E.E. Haaksma, T. Fox, GRID/CPCA: a new computational tool to design selective ligands, J. Med. Chem. 43 (2000) 3033–3044.
[5] T. Naumann, H. Matter, Structural classification of protein kinases using 3D molecular interaction field analysis of their ligand binding sites: target family landscapes, J. Med. Chem. 45 (2002) 2366–2378.
[6] C.A. Reynolds, R.C. Wade, P.J. Goodford, Identifying targets for bioreductive agents: using GRID to predict selective binding regions of proteins, J. Mol. Graph. 7 (103–108) (1989) 100.
[7] A. Miranker, M. Karplus, Functionality maps of binding sites: a multiple copy simultaneous search method, Proteins 11 (1991) 29–34.
[8] R. Bitetti-Putzer, D. Joseph-McCarthy, J.M. Hogle, M. Karplus, Functional group placement in protein binding sites: a comparison of GRID and MCSS, J. Comput. Aided Mol. Des. 15 (2001) 935–960.
[9] M.L. Verdonk, J.C. Cole, R. Taylor, SuperStar: a knowledge-based approach for identifying interaction sites in proteins, J. Mol. Biol. 289 (1999) 1093–1108.
[10] H. Gohlke, G. Klebe, DrugScore meets CoMFA: adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein, J. Med. Chem. 45 (2002) 4153–4170.
[11] R.A. Laskowski, J.M. Thornton, C. Humblet, J. Singh, X-SITE: use of empirically derived atomic packing preferences to identify favourable interaction regions in the binding sites of proteins, J. Mol. Biol. 259 (1996) 175–201.
[12] I.J. Bruno, J.C. Cole, J.P. Lommerse, R.S. Rowland, R. Taylor, M.L. Verdonk, IsoStar: a library of information about nonbonded interactions, J. Comput. Aided Mol. Des. 11 (1997) 525–537.
[13] D.R. Boer, J. Kroon, J.C. Cole, B. Smith, M.L. Verdonk, SuperStar: comparison of CSD and PDB-based interaction fields as a basis for the prediction of protein–ligand interactions, J. Mol. Biol. 312 (2001) 275–287.
[14] T. Lazaridis, M. Karplus, Effective energy functions for protein structure prediction, Curr. Opin. Struct. Biol. 10 (2000) 139–145.
[15] R.P. Sheridan, M.K. Holloway, G. McGaughey, R.T. Mosley, S.B. Singh, A simple method for visualizing the differences between related receptor sites, J. Mol. Graph. Model. 21 (2002) 217–225.
[16] Molecular Operating Environment (MOE) 2004.03, Chemical Computing Group Inc., 1010 Sherbrooke Street West, Suite 910, Montreal, Canada H3A 2R7, http://www.chemcomp.com.
[17] G. Ermondi, G. Caron, R. Lawrence, D. Longo, Docking studies on NSAID/COX-2 isoenzyme complexes using contact statistics analysis, J. Comput. Aided Mol. Des. 18 (2004) 683–696.
[18] G. Ermondi, M. Lorenti, G. Caron, Contribution of ionization and lipophilicity to drug binding to albumin: a preliminary step toward biodistribution prediction, J. Med. Chem. 47 (2004) 3949–3961.
[19] J. Baudry, W. Li, L. Pan, M.R. Berenbaum, M.A. Schuler, Molecular docking of substrates and inhibitors in the catalytic site of CYP6B1, an insect cytochrome p450 monooxygenase, Protein Eng. 16 (2003) 577–587.
[20] T. Kaminuma, Pathways and networks of nuclear receptors and modeling of syndrome X, Chem. Bio. Informatics J. 3 (2003) 130–156.
[21] B. Pirard, Peroxisome proliferator-activated receptors target family landscape: a chemometrical approach to ligand selectivity based on protein binding site analysis, J. Comput. Aided Mol. Des. 17 (2003) 785–796.
[22] R. Krishnan, E. Zhang, K. Hakansson, R.K. Arni, A. Tulinsky, M.S. Lim-Wilby, O.E. Levy, J.E. Semple, T.K. Brunck, Highly selective mechanism-based thrombin inhibitors: structures of thrombin and trypsin inhibited with rigid peptidyl aldehydes, Biochemistry 37 (1998) 12094–12103.
[23] L. Zhi, C.M. Tegley, B. Pio, J.P. Edwards, M. Motamedi, T.K. Jones, K.B. Marschke, D.E. Mais, B. Risek, W.T. Schrader, 5-Benzylidene-1,2-dihydrochromeno[3,4-f]quinolines as selective progesterone receptor modulators, J. Med. Chem. 46 (2003) 4104–4112.

[24] B.P. Morgan, A.G. Swick, D.M. Hargrove, J.A. LaFlamme, M.S. Moynihan, R.S. Carroll, K.A. Martin, E. Lee, D. Decosta, J. Bordner, Discovery of potent, nonsteroidal, and highly selective glucocorticoid receptor antagonists, J. Med. Chem. 45 (2002) 2417–2424.
[25] T.M. Sielecki, J.F. Boylan, P.A. Benfield, G.L. Trainor, Cyclin-dependent kinase inhibitors: useful targets in cell cycle regulation, J. Med. Chem. 43 (2000) 1–18.
[26] M.E. Noble, J.A. Endicott, L.N. Johnson, Protein kinase inhibitors: insights into drug design from structure, Science 303 (2004) 1800–1805.
[27] M.Y. Niv, H. Rubin, J. Cohen, L. Tsirulnikov, T. Licht, A. Peretzman-Shemer, E. Cna'an, A. Tartakovsky, I. Stein, S. Albeck, I. Weinstein, M. Goldenberg-Furmanov, D. Tobi, E. Cohen, M. Laster, S.A. Ben-Sasson, H. Reuveni, Sequence-based design of kinase inhibitors applicable for therapeutics and target identification, J. Biol. Chem. 279 (2004) 1242–1255.
[28] A. Levitzki, Protein kinase inhibitors as a therapeutic modality, Acc. Chem. Res. 36 (2003) 462–469.
[29] S.L. McGovern, B.K. Shoichet, Kinase inhibitors: not just for kinases anymore, J. Med. Chem. 46 (2003) 1478–1483.
[30] R.P. Sheridan, J. Shpungin, Calculating similarities between biological activities in the MDL drug data report database, J. Chem. Inf. Comput. Sci. 44 (2004) 727–740.
[31] N.R. Commitee, A unified nomenclature system for nuclear receptor subfamily, Cell 97 (1999) 1–20.
[32] F. Horn, G. Vriend, F.E. Cohen, Collecting and harvesting biological data: the GPCRDB and NucleaRDB information systems, Nucleic Acids Res. 29 (2001) 346–349.
[33] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.
[34] Z. Deng, C. Chuaqui, J. Singh, Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein–ligand binding interactions, J. Med. Chem. 47 (2004) 337–344.
[35] U. Lessel, D. Schomburg, Similarities between protein 3-D structures, Protein Eng. 7 (1994) 1175–1187.
[36] D. Schomburg, J. Reichelt, BRAGI—a comprehensive protein modeling program system, J. Mol. Graph. 6 (1988) 161–165.
[37] M.B. Eisen, P.T. Spellman, P.O. Brown, D. Botstein, Cluster analysis and display of genome-wide expression patterns, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 14863–14868.
[38] D.H. Williams, T. Mitchell, Latest developments in crystallography and structure-based design of protein kinase inhibitors as drug candidates, Curr. Opin. Pharmacol. 2 (2002) 567–573.
[39] A.C. Pike, A.M. Brzozowski, R.E. Hubbard, T. Bonn, A.G. Thorsell, O. Engstrom, J. Ljunggren, J.A. Gustafsson, M. Carlquist, Structure of the ligand-binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist, EMBO J. 18 (1999) 4608–4618.
[40] H. Greschik, J.M. Wurtz, S. Sanglier, W. Bourguet, A. van Dorsselaer, D. Moras, J.P. Renaud, Structural and functional evidence for ligand-independent transcriptional activation by the estrogen-related receptor 3, Mol. Cell 9 (2002) 303–313.
[41] P.M. Matias, P. Donner, R. Coelho, M. Thomaz, C. Peixoto, S. Macedo, N. Otto, S. Joschko, P. Scholz, A. Wegg, S. Basler, M. Schafer, U. Egner, M.A. Carrondo, Structural evidence for ligand specificity in the binding domain of the human androgen receptor. Implications for pathogenic gene mutations, J. Biol. Chem. 275 (2000) 26164–26171.
[42] P.M. Matias, M.A. Carrondo, R. Coelho, M. Thomaz, X.Y. Zhao, A. Wegg, K. Crusius, U. Egner, P. Donner, Structural basis for the glucocorticoid response in a mutant human androgen receptor (AR(ccr)) derived from an androgen-independent prostate cancer, J. Med. Chem. 45 (2002) 1439–1446.
[43] D. Rauh, G. Klebe, M.T. Stubbs, Understanding protein–ligand interactions: the price of protein flexibility, J. Mol. Biol. 335 (2004) 1325–1341.
[44] A.J. Tervo, T. Ronkko, T.H. Nyronen, A. Poso, BRUTUS: optimization of a grid-based similarity function for rigid-body molecular superposition. 1. Alignment and virtual screening applications, J. Med. Chem. 48 (2005) 4076–4086.
[45] M. Pastor, G. Cruciani, I. McLay, S. Pickett, S. Clementi, GRid-INdependent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors, J. Med. Chem. 43 (2000) 3233–3243.
[46] N. Stiefl, K. Baumann, Structure-based validation of the 3D-QSAR technique MaP, J. Chem. Inf. Model. 45 (2005) 739–749.
[47] R. Morphy, C. Kay, Z. Rankovic, From magic bullets to designed multiple ligands, Drug Discov. Today 9 (2004) 641–651.