Design and Quantitative Structure−Activity Relationship of 3-Amidinobenzyl-1*H*-indole-2-carboxamides as Potent, Nonchiral, and Selective Inhibitors of Blood Coagulation Factor Xa

Hans Matter,* Elisabeth Defossa, Uwe Heinelt, Peter-Michael Blohm, Detlev Schneider, Andrea Müller, Silke Herok, Herman Schreuder, Alexander Liesum, Volker Brachvogel, Petra Lönze, Armin Walser, Fahad Al-Obeidi, and Peter Wildgoose

Aventis Pharma Deutschland GmbH, DI&A, Molecular Modeling, Medicinal Chemistry, Structural Biology, DG Thrombosis and Degenerative Joint Diseases, Building G 878, D-65926 Frankfurt am Main, Germany

A series of 138 nonchiral 3-amidinobenzyl-1*H*-indole-2-carboxamides and analogues as inhibitors of the blood coagulation enzyme factor Xa (fXa) were designed, synthesized, and investigated by X-ray structure analysis and 3D quantitative structure−activity relationship (QSAR) studies (CoMFA, CoMSIA) in order to identify important protein−ligand interactions responsible for biological affinity and selectivity. Several compounds from this series are highly potent and selective inhibitors of this important enzyme linking extrinsic and intrinsic coagulation pathways. To rationalize biological affinity and to provide guidelines for further design, all compounds were docked into the factor Xa binding site. Those docking studies were based on X-ray structures of factor Xa in complex with literature-known inhibitors. It was possible to validate those binding modes by four X-ray crystal structures of representative ligands in factor Xa, while one ligand was additionally crystallized in trypsin to rationalize requirements for selective factor Xa inhibition. The 3D-QSAR models based on a superposition rule derived from these docking studies were validated using conventional and cross-validated r^2 values using the leave-one-out method and repeated analyses using two randomly chosen cross-validation groups plus randomization of biological activities. This led to consistent and highly predictive 3D-QSAR models with good correlation coefficients for both CoMFA and CoMSIA, which were found to correspond to experimentally determined factor Xa binding site topology in terms of steric, electrostatic, and hydrophobic complementarity. Subsets selected as smaller training sets using 2D fingerprints and maximum dissimilarity methods resulted in 3D-QSAR models with remarkable correlation coefficients and a high predictive power. The final quantitative SAR information agrees with all experimental data for the binding topology and thus provides reasonable activity predictions for novel factor Xa inhibitors.

1. Introduction

There is a growing interest in new, orally active anticoagulants with significant advantages to current agents such as heparin and warfarin for the treatment and prevention of thrombotic diseases. Those diseases involving deep vein thrombosis and stroke are one of the major reasons for mortality in Europe and the U.S.[1] The blood coagulation serine protease factor Xa (fXa) plays a central role in the coagulation cascade,[2] linking the intrinsic (activated by surface contact) and extrinsic (activated by vessel injury) pathways to the common coagulation pathway.[3] Factor Xa in combination with factor Va activates prothrombin on a phospholipid surface to generate thrombin,[4] while it is not known to be involved in processes other than hemostasis. Thrombin subsequently converts fibrinogen to fibrin, inducing clot formation and platelet aggregation.[5] Both incidents lead to serious pathological situations. The inhibition of fXa compared to thrombin may allow the effective control of thrombogenesis with minimal effect upon bleeding[6−10] because fXa inhibitors should affect coagulation specifically. Furthermore, inhibition of fXa is seen

to be more efficacious because one molecule fXa generates many thrombin molecules. In addition, inhibition of fXa should prevent new production of thrombin, while a basal thrombin level might be necessary for primary hemostasis. A growing number of reviews[11−15] and publications[16] demonstrate the pharmaceutical interest in potent and selective fXa inhibitors.

The mature form of factor X consists of a 139-residue light chain and a 303-residue heavy chain linked by a disulfide. It is synthesized in the liver and secreted after post-translational modifications into the blood as zymogen. Factor X is activated by the factor VIIa−tissue factor complex in the extrinsic pathway, initiated by vascular damage, or by the factor IXa−factor VIIIa complex in the intrinsic pathway. The heavy chain encompasses a serine protease domain in a trypsin-like closed *β*-barrel fold with the active triad Ser195-His57-Asp102 and two neighboring protein subsites S1 and S4, which are typically explored for obtaining high-affinity inhibitors.

In the first approach, peptidic factor Xa inhibitors were identified by high-throughput screening of combinatorial peptide libraries, resulting in selective and potent inhibitors such as SEL2711 (K_i(FXa) = 3 nM;


* To whom correspondence should be addressed. Phone: ++49-69-305-84329. Fax: ++49-69-331399. E-mail: hans.matter@aventis.com.

K_i(thrombin) = 40 000 nM[17]). Those structures allowed extraction of relevant structural motifs for factor Xa inhibition. The subsequent availability of high-resolution factor Xa X-ray structures in uncomplexed form[18] and with bound inhibitors[19] led us to the design and systematic synthesis of novel, achiral fXa inhibitors by a combination of benzamidines directed toward the fXa S1 pocket and basic or hydrophobic substituents directed toward the flexible S4 subsite of factor Xa. Because peptide inhibitors have insufficient pharmacokinetic properties such as low oral bioavailability and short half-lives after administration to animals, replacement by a rigid central scaffold as the priviledged structure was desirable. Suitable central scaffolds are carbocyclic or heterocyclic rings with possibilities for directing substituents toward the fXa S4 subpocket. This led to the identification of the 3-amidinobenzyl-1H-indole-2-carboxamide scaffold[20] as an interesting structural motif offering potency and selectivity by connecting benzamidines in order to provide a geometrical framework for orienting side chains in S4. Because of its simplicity, lack of chirality, ease of synthesis, and opportunity for structure-based design, this scaffold was chosen for systematic exploration. In this report, we combine structure-based and 3D quantitative structure–activity relationship (QSAR) approaches to understand relevant protein–ligand interactions for activity and selectivity of 138 3-amidinobenzyl-1H-indole-2-carboxamides and analogues. This is achieved by a combination of X-ray structure analysis in factor Xa and trypsin, flexible molecular docking, and 3D-QSAR analyses. The reported 3D-QSAR models allow reliable predictions of novel candidates and uncover binding features that are responsible for affinity and selectivity.

Although cationic interactions[21] in both S1 and S4 subsites are favorable for in vitro affinity, they might be detrimental for oral bioavailability. Hence, the replacement of basic moieties in the S4 pocket by lipophilic substituents was important. The use of 3D-QSAR models led to the identification of binding site regions, where steric, electronic, or hydrophobic effects play a dominant role. At the final stage, a solid-phase optimization strategy[22] was used that allowed replacement of the basic functionality in S4 toward lead optimization. Similar analyses were done for thrombin affinity in this dataset and helped to understand and improve the critical structural factors for selectivity.

The determination of the active conformation is the crucial step in 3D-QSAR. With the availability of protein 3D structures, this active conformation can be obtained by two approaches: docking of a reference compound and X-ray crystallography. The first approach was chosen[23] and could be validated by subsequent X-ray structure analyses of four representative compounds in factor Xa and one compound in the related serine protease trypsin. Flexible docking was based on available X-ray structures of non-peptidic benzamidine-type factor Xa inhibitors in complex with factor Xa.[19b,c,24] This led to an alignment for all other compounds by superimposing them onto the template and relaxing them within the cavity. This superposition produced consistent 3D-QSAR models that explained the key ligand–enzyme interactions.

Scheme 1. Relevant Protein−Ligand Interactions of 3-Amidinobenzyl-1H-indole-2-carboxamides in Complex with Human Factor Xa
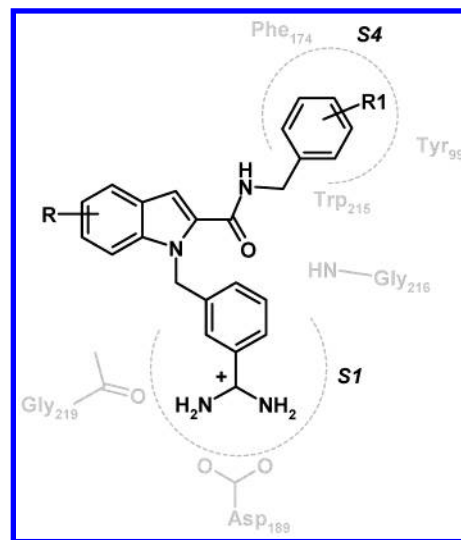
Comparative molecular field analysis (CoMFA)[25−27] and comparative molecular similarity index analysis (CoMSIA)[28] are used to derive relationships between molecular property fields and biological activities. Electrostatic and steric interaction energies are computed between each ligand and a probe atom located on predefined grid points for CoMFA, while for CoMSIA, those interaction fields are replaced by fields based on similarity indices between probe atoms and each molecule. The PLS method (partial least squares)[29] is used to derive a linear relationship for highly underdetermined matrices, and cross-validation[30] is used to check for consistency and predictiveness. The contour maps from 3D-QSAR models enhance the understanding of electrostatic, hydrophobic, and steric requirements for ligand binding, guiding the design of novel inhibitors to those regions where structural variations altering steric or electrostatic fields reveal a significant correlation to biological properties.

2. Methods

2.1. Design of Factor Xa Inhibitors. During our fXa project, the 3-amidinobenzyl-1H-indole-2-carboxamide scaffold was identified to lead to potent and selective inhibitors. We were successful using 2-substituted 3-amidinobenzyl-1H-indoles to orient different residues in the S4 subpocket (Scheme 1). Starting from N-methylpyridinium (**1**, **2**) and trimethyl-ammonium benzyl substitutions at the 3-amidinobenzyl-1H-indole-2-carboxamide scaffold (**76**, **77**), the best position for this positive charge directed toward S4 was investigated, followed by aliphatic replacement to improve their pharmacokinetic properties. Our design is based on the fXa active site topology from X-ray structures of des-Gla human factor Xa with known inhibitors[19] and some 3-amidinobenzyl-1H-indole-2-carboxamides. In agreement with other proteases (e.g., trypsin), the benzamidine moiety attached to the indole N1 interacts with Asp189 in the fXa S1 pocket.

3-Amidinobenzyl-1H-indole-2-carboxamides with polar groups directed toward the S4 subpocket are potent fXa inhibitors. Substitutions at positions 5 and 6 at the 3-amidinobenzyl-1H-indole-2-carboxamide scaffold reduce activity, while halogens in position 3 enhance activity. Small substituents such as hydroxy and methyl are preferred for position 4. The 3-amidinobenzyl group attached to indole N1 is essential for inhibition. The amide substituent at position 2 is directed toward the S4 pocket. Highly polar groups in this area can be

Table 1. 138 3-Amidinobenzyl-1*H*-Indole-2-carboxamides and Analogues Used in this Study

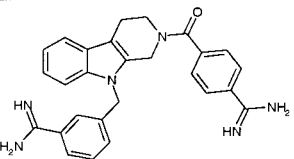| Compd	X	R^1	R^2	R^3	R^4	R^5	R^6	Ki (FXa) [nM]	Ki (Thr) [nM]
1	N-CH$_2$COOCH$_3$	(methylpyridinium-4-yl)-methyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	H	H	170	
2	N-ethyl	(methylpyridinium-4-yl)-methyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	H	H	500	
3	NH	3,5-dichlorobenzyl	3-amidinobenzyl	Br	H	H	H	85	
4	NH	3,5-dimethylbenzyl	3-amidinobenzyl	Br	H	H	H	13	
5	NH	3-amidinobenzyl	3-amidinobenzyl	Br	CH$_3$	H	H	50	2070
6	NH	3-amidinobenzyl	3-amidinobenzyl	Br	H	H	H	26	3290
7	NH	(methylpyridinium-4-yl)-methyl trifluoro acetic acid salt	3-amidinobenzyl	CH$_3$	H	H	H	170	
8	NH	1-naphthylmethyl	3-amidinobenzyl	Cl	H	H	H	304	
9	NH	3-amidinobenzyl	3-amidinobenzyl	Cl	H	H	H	9	4240
10	NH	4-(trimethylammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	Cl	H	H	H	40	
11	NH	(methylpyridinium-4-yl)-methyl trifluoro acetic acid salt	3-amidinobenzyl	COOCH$_3$	H	H	H	350	
12	NH	(methylpyridinium-4-yl)-ethyl trifluoro acetic acid salt	3-amidinobenzyl	H	OCH$_3$	H	H	160	
13	NH	(methylpyridinium-4-yl)-methyl trifluoro acetic acid salt	4-amidinobenzyl	H	H	H	H	600	
14	NH	(methylpyridinium-4-yl)-methyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	F	H	360	
15	NH	(methylpyridinium-4-yl)-methyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	H	H	150	
16	NH	(methylpyridinium-4-yl)-methyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	NH$_2$	H	100	57000
17	NH	(methylpyridinium-4-yl)-phenyl-methyl trifluoro acetic acid salt	3-amidinobenzyl	H	OCH$_3$	H	H	110	
18	NH	1-(S)-(methylpyridinium-4-yl)-ethyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	H	H	92	
19	NH	1-(S)-phenyl-ethyl trifluoro acetic acid salt	3-thiocarbamoyl-benzyl	H	H	OBn	H	4830	
20	NH	1-(S)-phenyl-ethyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	OBn	H	3690	
21	NH	1-amidino-piperidin-4-ylmethyl	3-amidinobenzyl	H	CH$_3$	H	H	25	
22	NH	1-methyl-quinolin-4-ylmethyl trifluoro acetic acid salt	3-amidinobenzyl	H	OH	H	H	50	
23	NH	1-naphthylmethyl	3-amidinobenzyl	H	CH$_3$	H	H	96	
24	NH	1-naphthyl-methyl	3-amidinobenzyl	H	OH	H	H	68	
25	NH	1-oxy-pyridin-4-yl-methyl trifluoro acetic acid salt	3-amidinobenzyl	H	CH$_3$	H	H	14480	
26	NH	2-(3,4-dichlorophenyl)-ethyl	3-amidinobenzyl	H	OH	H	H	918	
27	NH	2-(3-chlorophenyl)-ethyl	3-amidinobenzyl	H	OH	H	H	707	
28	NH	2-(4-bromophenyl)-ethyl	3-amidinobenzyl	H	OH	H	H	1628	
29	NH	2-(4-chlorophenyl)-ethyl	3-amidinobenzyl	H	OH	H	H	1556	
30	NH	2-(4-hydroxyphenyl)-ethyl	3-amidinobenzyl	H	OCH$_3$	H	H	970	
31	NH	2-(methyl-naphthalen-1-yl-amino)-ethyl	3-amidinobenzyl	H	OH	H	H	2470	
32	NH	2-(methylpyridinium-4-yl)-ethyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	H	H	1300	
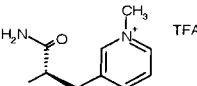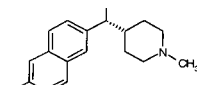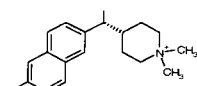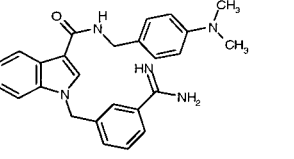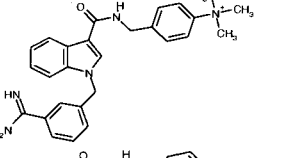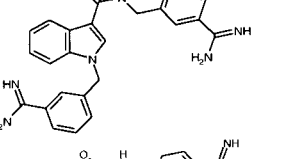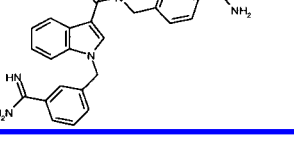
Table 1 (Continued)

Compd	X	R^1	R^2	R^3	R^4	R^5	R^6	Ki (FXa) [nM]	Ki (Thr) [nM]
33	NH	2-(naphthalen-1-ylamino)-ethyl	3-amidinobenzyl	H	OH	H	H	318	
34	NH	2,3,5-trichlorobenzyl	3-amidinobenzyl	H	OH	H	H	99	270
35	NH	2,4-dichlorophenyl-ethyl	3-amidinobenzyl	H	OH	H	H	840	
36	NH	2-chlorobenzyl	3-amidinobenzyl	H	OH	H	H	1200	
37	NH	2-naphthylmethyl	3-amidinobenzyl	H	OH	H	H	1720	
38	NH	3,4-dichlorobenzyl	3-amidinobenzyl	H	OH	H	H	841	
39	NH	3,4-dimethoxybenzyl	3-amidinobenzyl	H	OH	H	H	17350	
40	NH	3,5-bistrifluorobenzyl	3-amidinobenzyl	H	OH	H	H	99	1180
41	NH	3,5-dichlorobenzyl	3-amidinobenzyl	H	CH_3	H	H	25	1610
42	NH	3,5-dichlorophenyl	3-amidinobenzyl	H	OH	H	H	798	
43	NH	3,5-difluorobenzyl	3-amidinobenzyl	H	OH	H	H	191	
44	NH	3,5-dimethoxybenzyl	3-amidinobenzyl	H	OH	H	H	193	
45	NH	3,5-dimethylbenzyl	3-amidinobenzyl	H	CH_3	H	H	9	1420
46	NH	3,5-dimethylbenzyl	3-amidinobenzyl	H	OH	H	H	11	620
47	NH	3-amidinobenzyl	3-amidinobenzyl	H	CH_3	H	H	7	2390
48	O	3-amidinobenzyl	3-amidinobenzyl	H	H	OBn	H	61	
49	NH	3-amidinobenzyl	3-amidinobenzyl	H	H	NH_2	H	51	9900
50	NH	3-amidinobenzyl	3-amidinobenzyl	H	OH	H	H	9	3600
51	NH	3-amidinobenzyl	3-amidinobenzyl	H	H	OH	H	77	
52	NH	3-amidinobenzyl	3-amidinobenzyl	H	H	OBn	H	133	
53	NH	3-amidinobenzyl	3-amidinobenzyl	H	H	H	H	19	
54	NH	3-amidinobenzyl	3-amidinobenzyl	H	H	H	H	26	a)
55	NH	3-bromobenzyl	3-amidinobenzyl	H	OH	H	H	117	
56	NH	3-carbamidoylbenzyl	3-amidinobenzyl	H	OH	H	H	180	
57	NH	3-carbamoylbenzyl	3-amidinobenzyl	H	H	H	H	240	
58	NH	3-chlorobenzyl	3-amidinobenzyl	H	OH	H	H	262	
59	NH	3-fluoro-5-trifluoromethyl-benzyl	3-amidinobenzyl	H	OH	H	H	39	1710
60	NH	3-fluorobenzyl	3-amidinobenzyl	H	OH	H	H	720	
61	NH	3-hydroxycarbonylbenzyl	3-amidinobenzyl	H	OH	H	H	10000	
62	NH	3-iodobenzyl	3-amidinobenzyl	H	OH	H	H	42	
63	NH	3-methoxybenzyl	3-amidinobenzyl	H	OH	H	H	308	
64	NH	3-methoxycarbonyl-benzyl	3-amidinobenzyl	H	OH	H	H	2043	
65	NH	3-methyl-5-trifluoro-benzyl	3-amidinobenzyl	H	OH	H	H	13	480
66	NH	3-methylbenzyl	3-amidinobenzyl	H	OH	H	H	131	
67	NH	3-nitrobenzyl	3-amidinobenzyl	H	CH_3	H	H	10000	
68	NH	3-trifluorobenzyl	3-amidinobenzyl	H	OH	H	H	476	
69	NH	4-(benzyl-dimethyl-ammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	CH_3	F	H	245	
70	NH	4-(benzyl-dimethyl-ammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	CH_3	H	H	23	3900
71	NH	4-(dimethylamino)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	CH_3	H	H	1310	
72	NH	4-(dimethylamino)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	F	H	3520	
73	NH	4-(dimethyl-ethyl-ammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	F	H	103	
74	NH	4-(dimethyl-prop-2-enyl-ammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	F	H	141	
75	NH	4-(dimethyl-prop-2-ynyl-ammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	F	H	76	
76	NH	4-(trimethylammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	H	H	90	
77	NH	4-(trimethylammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	H	OCH_3	1760	
78	NH	4-(trimethylammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	OCH_3	H	H	62	

Table 1 (Continued)

Compd	X	R^1	R^2	R^3	R^4	R^5	R^6	Ki (FXa) [nM]	Ki (Thr) [nM]
79	NH	4-(trimethylammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	OBn	H	82	
80	NH	4-(trimethylammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	F	H	66	
81	NH	4-(trimethylammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	Cl	H	140	15800
82	NH	4-(trimethylammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	OH	H	28	11900
83	NH	4-(trimethylammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	CH_3	H	H	14	
84	NH	4-(trimethylammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	NO2	H	144	
85	NH	4-(trimethylammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	NHBOC	H	14	28000
86	NH	4-(trimethylammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	SO_2CH_3	H	226	
87	NH	4-(trimethylammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	OH	H	H	7	9090
88	NH	4-(trimethylammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	OCH_3	H	94	
89	NH	4-(trimethylammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	NH_2	H	60	
90	NH	4-(trimethylammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	OBn	H	H	84	
91	NH	4-(trimethylammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	CH_3	H	H	16	b)
92	NH	4-(trimethylammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	Br	H	H	24	
93	NH	4-(trimethylammonium)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	Br	H	134	
94	NH	4-(trimethylammonium-methyl)-benzyl trifluoro acetic acid salt	3-amidinobenzyl	H	H	H	H	400	
95	NH	4-amidinobenzyl	3-amidinobenzyl	H	H	H	H	260	
96	NH	4-amidinobenzyl	3-amidinobenzyl	H	OCH_3	H	H	131	
97	NH	4-amidinobenzyl	3-amidinobenzyl	H	H	NH_2	H	89	5600
98	NH	4-amidinobenzyl	3-amidinobenzyl	H	OH	H	H	208	4200
99	NH	4-amidino-cyclohexyl-methyl	3-amidinobenzyl	H	CH_3	H	H	45	
100	NH	4-chlorobenzyl	3-amidinobenzyl	H	H	H	H	3600	
101	NH	4-hydroxybenzyl	3-amidinobenzyl	H	OCH_3	H	H	183	
102	NH	4-hydroxybenzyl	3-amidinobenzyl	H	CH_3	H	H	1608	
103	NH	4-methoxybenzyl	3-amidinobenzyl	H	OH	H	H	1360	
104	NH	4-pyridyl-methyl	3-amidinobenzyl	H	OCH_3	OCH_3	H	5310	
105	NH	adamantylmethyl	3-amidinobenzyl	H	OH	H	H	274	
106	NH	biphenyl-3-ylmethyl	3-amidinobenzyl	H	CH_3	H	H	325	
107	NH	Bn	3-amidinobenzyl	H	H	H	H	6890	
108	NH	Bn	3-amidinobenzyl	H	OH	H	H	2032	
109	NH	cyclohexylmethyl	3-amidinobenzyl	H	OH	H	H	1112	
110	NH	H	3-amidinobenzyl	H	H	H	H	5650	
111	NH	H	3-amidinobenzyl	H	CH_3	H	H	7910	
112	NH	H	3-amidinobenzyl	H	OH	H	H	7160	
113	NH	isoquinolin-1-yl-methyl	3-amidinobenzyl	H	CH_3	H	H	15	200
114	NH	isoquinolin-1-yl-methyl	3-amidinobenzyl	H	OH	H	H	20	290

Table 1 (Continued)

Compd	X	R^1	R^2	R^3	R^4	R^5	R^6	Ki (FXa) [nM]	Ki (Thr) [nM]
115	NH	naphthalen-2-lcarbamoyl-methyl	3-amidinobenzyl	H	OH	H	H	2384	
116	NH	phenyl-pyridin-4-yl-methyl	3-amidinobenzyl	H	OCH$_3$	H	H	6790	
117	NH	phenyl-pyridin-4-yl-methyl	3-amidinobenzyl	H	H	H	H	7030	
118	NH	phenylsulfonamidyl	3-amidinobenzyl	H	OH	H	H	7050	
119	NH	pyren-1-ylmethyl	3-amidinobenzyl	H	OH	H	H	2064	
120	NH	pyridin-2-ylmethyl	3-amidinobenzyl	H	OH	H	H	3327	
121	NH	quinolin-4-ylmethyl	3-amidinobenzyl	H	OH	H	H	300	
122	NH	quinolin-4-yl-methyl	3-amidinobenzyl	H	CH$_3$	H	H	173	
123	O	CH$_3$	3-amidinobenzyl	CH$_3$	H	H	H	8800	
124	O	CH$_3$	3-amidinobenzyl	COOCH$_2$	H	H	H	7500	
125	O	3-amidinobenzyl	3-amidinobenzyl	H	H	H	H	28	2460
126	O	3-amidinobenzyl	3-amidinobenzyl	H	CH$_3$	H	H	44	
127	O	C$_2$H$_5$	4-amidinobenzyl	H	H	H	H	55000	
128	O	C$_2$H$_5$	3-amidinobenzyl	H	H	H	H	4200	
129	O	C$_2$H$_5$	3-amidinobenzyl	H	CH$_3$	H	H	948	
130	O	CH$_3$	3-amidinobenzyl	H	H	F	H	13000	
131	O	H	3-amidinobenzyl	H	H	OH	H	10000	
132	S	CH$_3$	3-amidinobenzyl	H	OH	H	H	2986	
133	NH		3-amidinobenzyl	H	H	H	H	3900	
134	NH		3-amidinobenzyl	H	H	H	H	610	
135	NH		3-amidinobenzyl	H	H	H	H	1310	

Compd	Structure	Ki (fxa) [nM]
136		900
137		2190
138		1820
139		1230
140		1118

^{a} Diacetate salt of compound **53**, not used for 3D-QSAR. ^{b} Diacetate salt of compound **83**, not used for 3D-QSAR.

Scheme 2[a]
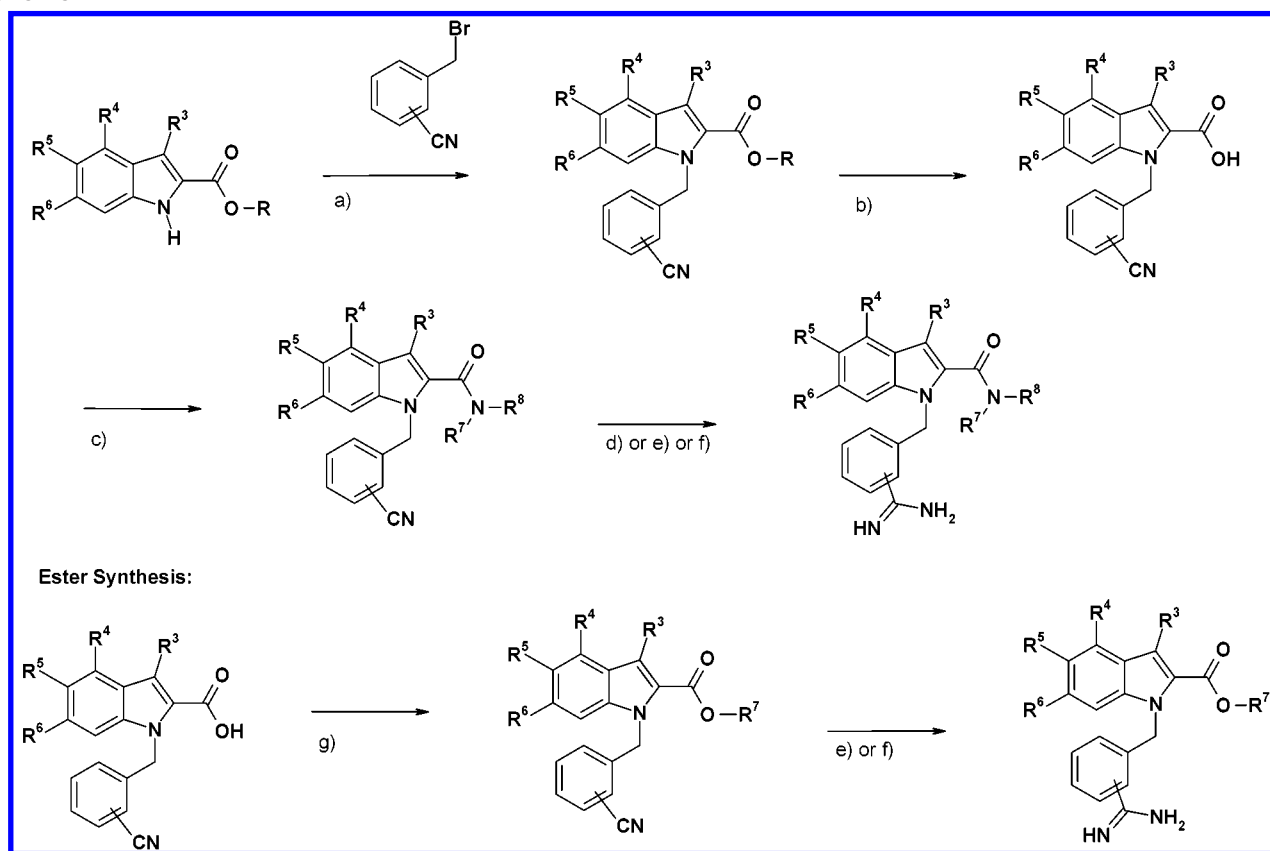
[a] (a) NaH, DMF; (b) NaOH, MeOH; (c) diphenylphosphoryl azide, DIPEA, DMF, HNR[7]R[8]; (d) (i) HCl, EtOH, (ii) NH$_3$, EtOH; (e) (i) HONH$_2$, EtOH, (ii) Raney nickel, NH$_4$Cl, EtOH/H$_2$O; (f) (i) H$_2$S, TEA, pyridine, (ii) MeI, acetone, (iii) NH$_4$OAc, HOAc, MeOH; (g) K$_2$CO$_3$, DMF, Br−R[7].

replaced by lipophilic substituents such as 3,5-dimethylbenzyl without loss of activity (**22**, **24**). This reduction of polarity is a step toward improving the pharmacokinetics. Direct connection of the 3,5-dichlorophenyl ring to the amide nitrogen results in a dramatic loss of activity. On the basis of these SAR considerations for 3-amidinobenzyl-1*H*-indole-2-carbox-amides supported by extensive modeling studies, novel compounds were planned and synthesized, as summarized in Table 1.

2.2. Chemistry and Enzyme Assay. Synthesis of the compounds listed in Table 1 was achieved by starting from commercially available or easily accessible indole 2-carboxylic acid esters in a conventional synthesis sequence.[20,22] Alkylation with 3-cyanobenzylbromide, ester cleavage, and coupling with the corresponding amines or alkylation with the corresponding bromides resulted in cyano precursors that were subsequently converted to amidines by known procedures such as the Pinner reaction (ref 31, not for indole esters), hydroxyamidine reduction,[32] and the H$_2$S/NH$_3$ method[33] respectively (Scheme 2). In addition, a solid-phase strategy (Scheme 3) was employed for the synthesis of compounds with hydroxy or amino functionality at C4 or C5 of the indole scaffold.[22] The biological assay used for compound testing was described earlier.[17a] For 3D-QSAR studies, enzyme inhibition is expressed as log[(1/K_i)100000]. Two compounds were synthesized and tested both as di-trifluoroacetate and di-acetate to assess the influence of the counterions on biological data, while only data from di-TFA salts were used for 3D-QSAR, resulting in 138 molecules for statistical analysis (**53/54** and **83/91** in Table 1).

2.3. X-ray Structure Analysis. 2.3.1. Crystallization. Purified human fXa was purchased from Enzyme Research Lab (South Bend, IN). The Gla domain was removed and the Gla-less factor Xa was crystallized in hanging drops as described earlier.[19] Trypsin from bovine pancreas (Sigma ref T8918) was dissolved in 0.6 M ammonium sulfate with 1 mg/mL CaCl$_2$ and 10 mg/mL benzamidine to a final concentration of 60 mg/mL. Crystals were grown at 19 °C using the hanging drop and vapor diffusion method and appeared in droplets equilibrated over reservoirs composed of 1.2−1.6 M ammonium sulfate adjusted with concentrated NaOH to pH 5.0−7.0.

2.3.2. Data Collection and Processing. Crystals were soaked in 5 μL of reservoir solution containing ∼20 mM or saturated inhibitor (depending on the solubility) for 24−72 h. The trypsin crystal was mounted in an X-ray capillary and measured at room temperature. The factor Xa crystals were measured at cryogenic temperatures. For cryogenic data collection, the crystals were picked up with a fiber loop, soaked for a few seconds in a solution containing 20% glycerol and 5−20 mM inhibitor in reservoir solution, and flash-frozen in a stream of gaseous nitrogen at 100 K.

The X-ray intensity data were collected on Mar300 imaging plate (trypsin and fXa-**79** and **125**) (X-ray research, Germany) or a Mar345 imaging plate (fXa-**41** and **45**), both mounted on an FR591 rotating anode (Nonius, The Netherlands) operating at 40 kV, 80 mA. Data processing and scaling were carried out using the XDS package.[34] Data collection and refinement statistics are presented in Table 2.

2.3.3. Structure Solution and Crystallographic Refinement. The structures were solved by molecular replacement. The search models were made from the coordinates of refined structures of trypsin and fXa complexes solved previously with the same crystal packing as the current complexes. The bound inhibitors were omitted from the search models. Energy-restrained least-squares refinement was carried out using X-PLOR.[35] Refinement was started with rigid-body refinement to adjust for small differences in cell dimensions, followed by energy minimization and individual temperature factor refinement. At this stage, the $2F_o - F_c$ and $F_o - F_c$ maps were inspected and the inhibitors were fitted. Solvent molecules were included if they were on sites of different electron density with values above 3.5σ and if they were within 3.5 Å of the protein molecule or a water molecule. After two to three

Scheme 3[a]
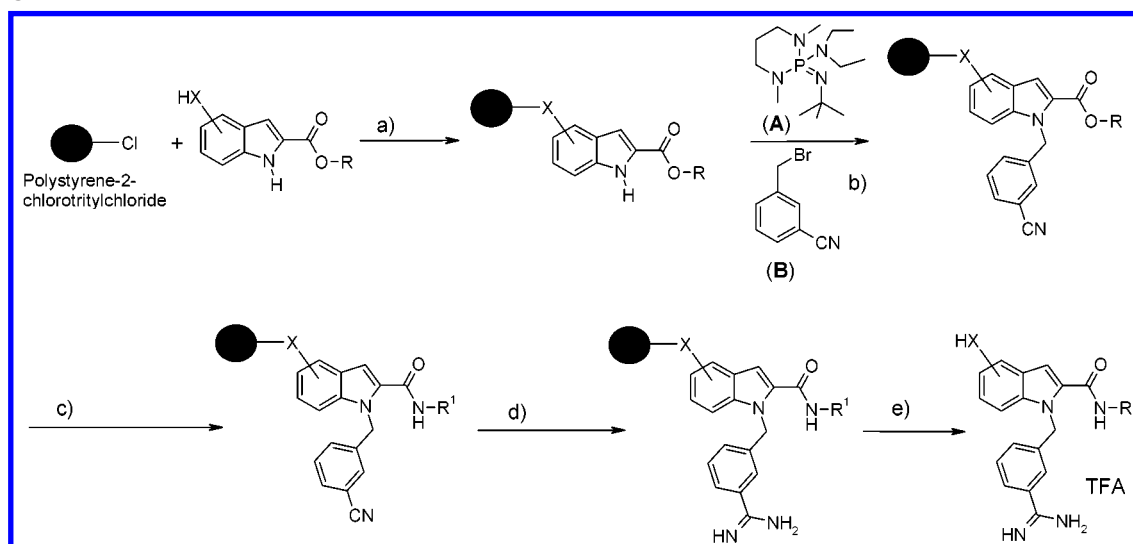
[a] (a) DCM/THF, DIPEA; (b) **A**, **B**, DMF; (c) (i) benzyltrimethylammonium hydroxide, MeOH, DMF, (ii) H_2N-R^1, DMF, DIC, HOBt (DIPEA); (d) (i) H_2S, pyridine, TEA, (ii) MeI, acetone, (iii) ammonium acetate, HOAc, MeOH; (e) DCM, TFA, H_2O.

Table 2. Data Collection and Refinement Statistics

enzyme	trypsin	fXa	fXa	fXa	fXa
inhibitor	**79**	**79**	**125**	**41**	**45**
space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
a, Å	63.6	56.2	56.7	56.3	56.7
b, Å	63.6	72.0	72.2	72.0	72.4
c, Å	68.9	78.7	78.6	78.3	78.5
obsd reflns	45 455	70 257	37 057	45 424	23 354
unique reflns	14 008	20 233	15 227	12 329	8 383
resolution, Å	2.2	2.0	2.2	2.4	2.7
R_{sym}	9.8	5.9	6.5	6.7	7.8
completeness, %	95.2	91.1	89.8	95.2	90.1
protein atoms	1629	2249	2249	2249	2249
inhibitor atoms	33[a]	41	64/32[b]	32	32
calcium ions	1	1	1	1	1
sulfate ions	1	0	0	0	0
water molecules	139	295	296	291	294
R_{factor}, %	16.6	19.7	19.2	18.1	16.1
R_{free}, %		28.6	28.0	28.0	29.5
rmsd bond length, Å	0.008	0.008	0.008	0.008	0.007
rmsd bond angle, deg	1.59	1.42	1.49	1.46	1.38
rmsd dihedral, deg	16.4	24.2	24.7	24.7	24.7
rmsd improper, deg	1.35	1.87	2.07	1.90	1.96

[a] The *O*-benzyl group of the inhibitor is not fitted because it is not visible in the electron density maps. [b] The inhibitor is bound in two different conformations with almost equal occupancy.

additional rounds of manual inspection, rebuilding, and refinement, the final models were obtained with *R* factors between 16.1% and 19.7% and free *R* factors between 28.6% and 29.5% and excellent geometry. The EGF-1 domain is not visible in the electron density maps probably because of disorder, and the rather high free *R* factors might relate to this disordered EGF-1 domain. The statistics of the crystallographic refinement are listed in Table 2.

2.4. Computational Procedures. 2.4.1. General Procedures. All modeling work and statistical analyses were performed using the program SYBYL[36] on Silicon Graphics workstations. Docking and QSAR model analyses were automated using scripts in PERL and SPL (Sybyl Programming Language). All energy calculations were based on the MMFF94s force field[37] using MMFF94 charges.[37] Conformations of ligands and complexes were minimized using quasi-Newton−Raphson (BFGS) or conjugate gradient (CG) procedures.

2.4.2. Docking Studies. For flexible docking, the factor Xa crystal structure 1XKA[19b] was used, while later investigations were based on the structures for **41**, **45**, **79**, and **125**. After analysis of key protein−ligand interactions using the program GRID,[38] candidate molecules were manually docked in differ-

ent orientations in the active site. Subsequently the resulting protein−ligand complexes were minimized, treating all ligand atoms plus protein residues within a sphere of 5 Å around the ligand as flexible, while the remaining protein was only used to compute nonbonded interactions. A structurally conserved water consistently situated in the S4 pocket was included in all energy minimizations. Other compounds were built accordingly, docked into fXa, and minimized. This alignment for all molecules in Table 1 served as the basis for 3D-QSAR. Automatic docking of selected compounds using the programs QXP[39] or FlexX[40] led to similar conformations, which were validated by X-ray structures in human des-Gla factor Xa. In total, X-ray structures for four 3-amidinobenzyl-1*H*-indole-2-carboxamides in factor Xa (**41**, **45**, **79**, and **125**) and one in trypsin (**79**) were obtained with resolutions ranging from 2.0 to 2.7 Å.

2.4.3. 3D-QSAR. Steric and electrostatic interaction energies between a probe atom and aligned molecules are calculated at predefined grid points with 2 Å spacing, a positively charged carbon atom, and a distance-dependent dielectric constant. Default settings were used for CoMFA and CoMSIA, if not stated otherwise.

The alignment was also used to compute steric, electrostatic, and hydrophobic similarity index fields for CoMSIA.[28] The advantage of CoMSIA is that no singularities occur at atomic positions because of a Gaussian-type distance dependence of the physicochemical properties. Similarity indices[41] were computed using a probe with a charge of +1, a radius of +1, a hydrophobicity of +1, and an attenuation factor α of 0.3 for Gaussian-type distance dependence.

Equal weights for CoMFA or CoMSIA fields were assigned using the CoMFA-STD scaling option.[42] Cross-validated analyses were run using the leave-one-out method in SAMPLS[43] or two and five cross-validation groups with a random selection of group members. PLS analyses using two or five randomly selected cross-validation groups were averaged over 100 runs. CoMFA columns with a variance smaller than 2.0 were excluded prior to the PLS analysis (minimum σ). The overall quality of all PLS analyses was expressed using the cross-validated value $r^2(cv)$. For validation, all biological activities were randomized[44] 100 times and were subjected to PLS analysis and the mean cross-validated r^2 was calculated.

2.4.4. 3D-QSAR Model Validation. Because simple randomization of activities could be misleading, if there is appreciable redundancy in the dataset, progressive shuffling[45] was used as an alternative randomization technique. This y-block perturbation strategy first subdivided biological activities into 2−12 subgroups of similar ranges. Subsequently the biological activities were randomized only within a subgroup,

while the relation between individual subgroups remained unaffected. Thus, the stability of each obtained regression model could be directly tested. For 12 subgroups, only a small local perturbation of biological activities was probed, while for 2 subgroups, a much larger portion of the dataset is randomized. This method provides information about the structure of the y block, indicates inconsistencies, and helps to estimate the tolerable error in biological activities for any PLS model. For analysis, the number of subgroups was plotted on the *x* axis against the mean cross-validated r^2 value per submodel (20 randomizations each) on the *y* axis.

Two-dimensional fingerprints were generated using the program UNITY[46] for rational selection of multiple training and test sets in order to evaluate the predictive capabilities of the final PLS model, which contains information about the presence of molecular fragments in a binary format. Their similarity is computed using the Tanimoto coefficient,[47] defined by the number of bit sets in both bit strings normalized by the number of bit sets in common. Compound selections were done using *maximum dissimilarity* methods and 2D fingerprints.[48,49]

3. Results and Discussion

3.1. X-ray Structure Analysis. Crystals of the factor Xa−inhibitor complexes diffracted well with resolution between 2.0 Å (**79**) and 2.7 Å (**45**) (Table 2), and the inhibitors generally were well defined, as seen in Figure 1. The crystal of the trypsin−**79** complex diffracted to 2.2 Å, and the inhibitor, with the exception of the *O*-benzyl moiety, was well defined (Figure 1E). The structures of the factor Xa complexes are similar to the structures of other factor Xa complexes in the protein databank. Superposition of the Cα atoms of the catalytic domain (residues 16−244) of the **79** complex gave the following root-mean-square deviations: 0.3 Å/234 Cα atoms for PDB entries 1F0R and 1FJS, 0.5 Å/234 Cα atoms for PDB entry 1XKA, and 0.6 Å/229 Cα atoms for PDB entry 1HCG. These root-mean-square differences are more influenced by the crystal packing (1XKA and 1HCG have a different crystal packing) than by the inhibitor.

3.2. Binding Mode of Indole-Based fXa Inhibitors. In all investigated fXa−inhibitor complexes, the 2-amidinobenzyl at the indole N1 is located in the fXa S1 pocket and interacts with Asp189 at the bottom of this pocket. This is the typical binding mode observed for other benzamidine-based factor Xa inhibitors. The indole scaffold itself is solvent-exposed and stacked upon the flexible side chain of Gln192, while it is also involved in van der Waals contacts with the Cys191−Cys220 disulfide bond.

The inhibitor **79** (fXa, 82 nM) carries an additional *O*-benzyl group at indole position 5, which is not visible in the trypsin−**79** complex (Figure 1E) in the electron density maps probably because of disorder. However, this group is clearly visible in the fXa−**79** complex and stacked against the Arg143 side chain. To accommodate this substituent, some reorganization of protein side chains takes place. While the Arg143 guanidine moiety rotates about 90°, the neighboring Gln192 side chain is also shifted, causing the formation of an additional hydrogen bond between Gln192-OE1 and Arg143-NH1.

The amide oxygens of **41** (fXa, 25 nM), **45** (fXa, 9 nM), and **79** are involved in hydrogen bonding interactions via a structurally conserved water to Ser214 C=O and the side chains of the catalytic triad residues His57 and Ser195. In contrast, the ester carbonyl oxygen of **125**

(fXa, 28 nM) is involved in a longer (∼3.3 Å) hydrogen bond with Gly216-N.

The terminal groups of the 3-amidinobenzyl-1*H*-indole-2-carboxamides and esters are situated in the fXa S4 pocket, lined by residues Tyr99, Phe174, Trp215, and Glu217. Since the main differences between the inhibitors occur here, the positions of these groups are discussed separately for each inhibitor.

For compound **41**, the terminal dichlorobenzyl substituent is oriented almost parallel to the aromatic side chain of Trp215 at the bottom of S4 and perpendicular to the rings of Tyr99 and Phe174. The two Cl atoms attached to the benzyl ring in this pocket have strong electron density (Figure 1). Cl-1 points deep into the S4 pocket, while Cl-2 points toward the solvent and in the direction of Glu217. The binding mode of inhibitor **45** is very similar because two chlorine atoms are replaced by methyl groups. Interestingly, both X-ray structures reveal that hydrophobic groups such as chlorine and methyl are located at positions that are known to preferably bind positively charged groups in the fXa S4 pocket. Thus, we could demonstrate that it is possible to replace basic and/or polar substituents by hydrophobic residues and still maintain low nanomolar fXa binding affinity.

The S4-directed 3-amidinobenzyl group in the fXa−**125** complex is bound in two different conformations. In one binding mode, the S4 amidine group is located at the same position as Cl-1 in the fXa−**41** complex, while in the other conformation, this amidine group occupies the position of Cl-2 in the fXa−**41** complex (Figure 2). The binding mode of **125** is otherwise identical to those of **41** and **45**. The 4-(trimethylammonium)benzyl group of compound **79** is the only para-substituted terminal group of the inhibitors crystallized for this study. The aromatic ring in **79** is bound almost perpendicular to the aromatic terminal rings of the meta-substituted inhibitors (Figure 2). It is bound almost parallel to the aromatic ring of Tyr99 and perpendicular to the Trp215 indole ring, while the charged center is superimposed over all other side chains (Figure 2). As discussed below, the position of the terminal ring of this inhibitor in factor Xa is similar to the position of this ring in the trypsin−**79** complex.

To obtain an alignment rule, the two reference inhibitors **41** and **45** were manually docked into factor Xa and optimized using the MMFF94S force field while treating all protein side chains within a sphere of 5 Å as flexible. All other compounds were treated similarly to obtain a superposition and hypothetical binding modes for all of them suitable for 3D-QSAR studies.

3.3. CoMFA Model and Validation Studies. All obtained 3D-QSAR models based on the alignment rule derived from the fXa active site topology showed a high degree of internal consistency. This section summarizes the statistical results for both CoMFA and CoMSIA models and the model validation studies. When a 2 Å grid spacing was used, a CoMFA model with an r^2(cv) value of 0.760 for five relevant PLS components and a conventional r^2 of 0.913 was obtained (model A, cf. Table 3 and Figure 3). The statistical number of PLS components in general is higher than the number of components required for a meaningful chemical interpretation. In order not to produce overdimensioned PLS models,
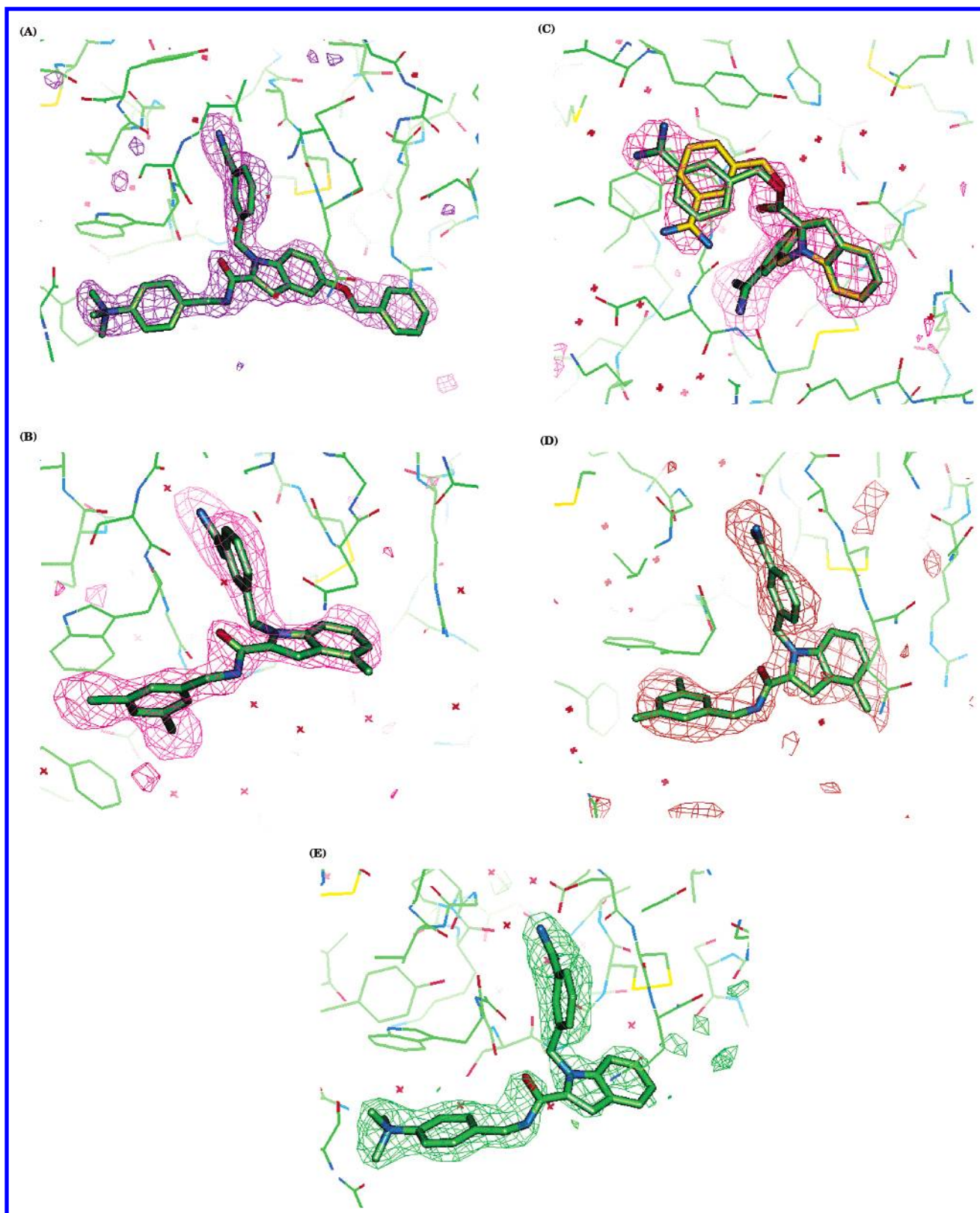
Figure 1. Comparison of X-ray binding modes for selected 3-amidinobenzyl-1*H*-indole-2-carboxamides in complex with human factor Xa and trypsin. Molecules (A) **79** (82 nM), (B) **41** (25 nM), (C) **125** (28 nM, two conformers), and (D) **45** (9 nM) are shown in complex with factor Xa, and (E) molecule **79** is shown in complex with bovine trypsin (1507 nM). The electron densities shown are from $F_o - F_c$ omit maps contoured at $(2.5-3.0)\sigma$. The inhibitors are well defined in the electron density maps with exception of the *O*-benzyl group and part of the indole ring of **79** in trypsin, which is probably a result of the nonoptimal fit of this inhibitor to the trypsin active site.

a component was seen as "optimal", if the q^2 is not significantly increased by adding an additional component to the model.

The alignment based on experimental information on the fXa binding site topology and the graph of observed versus fitted biological activities[50] for this model are
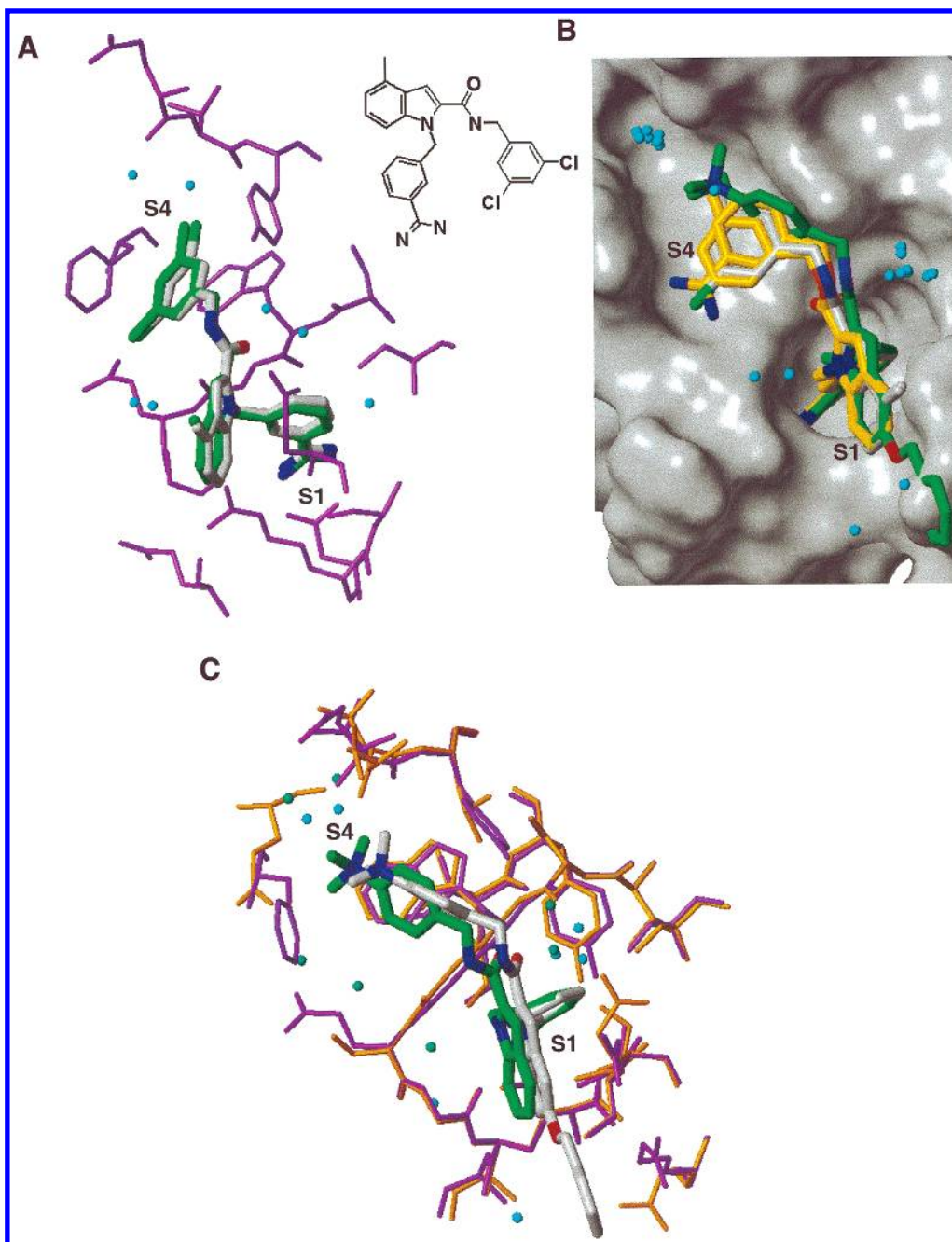
Figure 2. (A) Superposition of the best manual docking mode (green) of molecule **41** (25 nM) with the experimentally obtained complex (white ligand carbon atoms) in human factor Xa from X-ray structure analysis. Only selected residues from the fXa binding site are shown, conserved water residues are given as spheres, and all hydrogens are omitted for clarity. (B) Superposition of experimentally obtained X-ray structures of 3-amidinobenzyl-1*H*-indole-2-carboxamides (**41**, **45**, **79**, and **125**) with solvent-accessible surface for the fXa binding site generated from the fXa−**41** complex. Hydrogens are omitted for clarity. (C) Superposition of the crystal structures of **79** in fXa (purple/white) and trypsin (orange/green). The *O*-benzyl group at the indole ring is not present in the trypsin structure because it was not visible in the electron density maps. The inhibitor binding modes are similar, indicating that trypsin is a good surrogate for fXa here and correctly predicted the binding mode of **79**. However, as discussed in the text, details of the binding and the binding affinities differ (K_i(fXa) = 82 nM; K_i(trypsin) = 1507 nM).

displayed in Figure 3. The steric field descriptors (2156 grid-based variables) explain 59% of the variance, while the electrostatic field accounts for 41%. This relative percentage of field contributions is obtained taking the normalized sum of standard deviations after CoMFA_STD scaling multiplied by the PLS coefficients for the final non-cross-validated PLS model. Affinity predictions obtained from this and subsequent models are listed in Table 4.

Changing the grid spacing for CoMFA steric and electrostatic fields to 1 Å (16 632 grid-based variables per field type) led to a five-component PLS model (model B) with a slightly lower r^2(cv) value of 0.753, a conventional r^2 of 0.912, and a similar chemical interpretation, when inspecting the CoMFA contour maps (see below). Model A with 2 Å grid spacing was extensively validated to assess its predictive power and significance.

The effect of the compound alignment relative to the grid position within the predefined region definition file was evaluated by consistently moving all compounds in increments of 0.5 Å in all three dimensions *x*, *y*, and *z*. Here, the relative alignment is not changed, but the
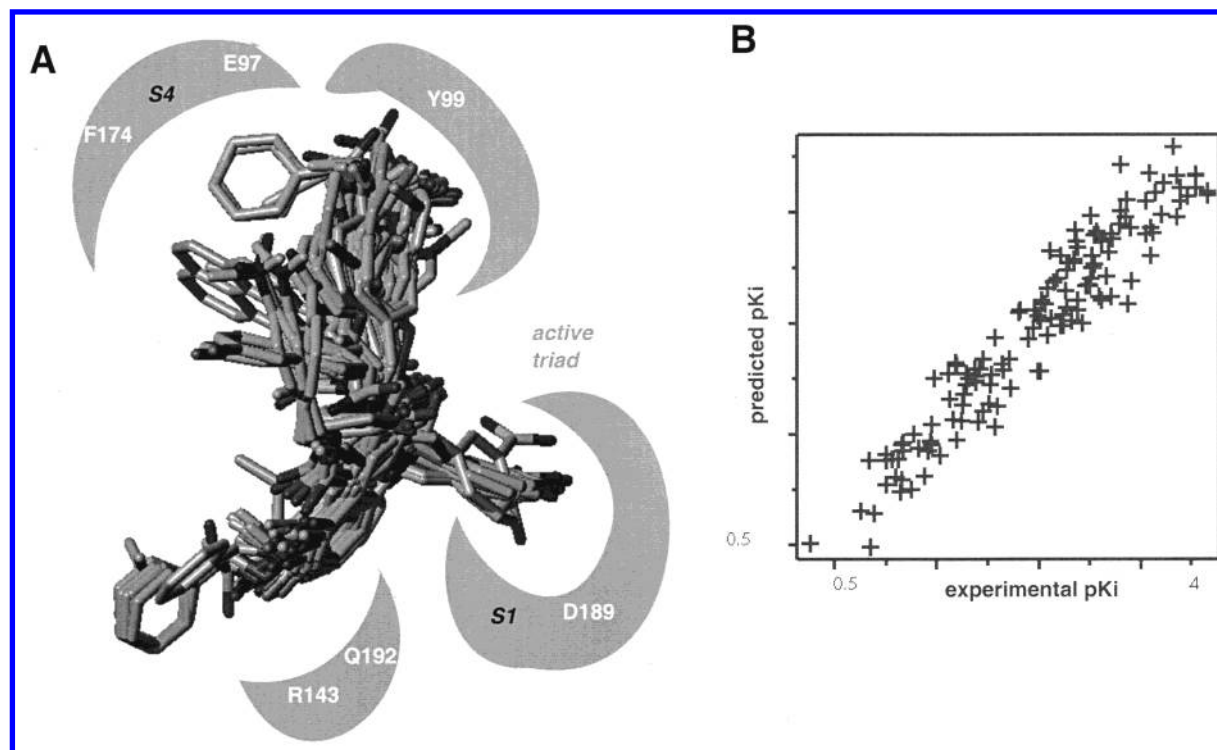
Figure 3. (A) Superposition of 138 3-amidinobenzyl-1*H*-indole-2-carboxamides and derivatives as factor Xa inhibitors for 3D-QSAR studies (cf. Table 1), built on the basis of the docked, *bioactive* conformation of the reference compounds **41**, **45**, **79**, and **125** and fitted into the protein cavity. Hydrogens are omitted for clarity. (B) Graph of observed versus fitted biological activity for the final CoMFA model with 2 Å grid spacing.

Table 3. Summary of 3D-QSAR Models To Explain Factor Xa Activity of 138 3-Amidinobenzyl-1*H*-indole-2-carboxamides and Derivatives[a]

	model	no. of compds	grid spacing [Å]	r^2(cv)	SD	PLS components	r^2
CoMFA[c]	A	138	2	0.760	0.450	5	0.913
CoMFA	B	138	1	0.753	0.457	5	0.912
CoMSIA[c]	C	138	2	0.655	0.540	5	0.863
CoMSIA	D	138	1	0.657	0.539	5	0.866
CoMFA	E	82[b]	2	0.698	0.518	4	0.938
CoMSIA	F	82[b]	2	0.660	0.539	5	0.933

[a] Two compounds indicated in Table 1 were not used for 3D-QSAR studies. All models were derived using a minimum σ of 2. r^2(cv): cross-validation r^2 using leave-one-out model. SD: standard deviation of error from leave-one-out PLS model. PLS components: optimal number of components. r^2: non-cross-validated regression coefficient. [b] Compounds used for this model are marked as SAR set 1 in Table 4. [c] Models A and C were subjected to extensive statistical validation.

absolute orientation with respect to the grid has changed. The r^2(cv) values for each orientation ranged from 0.712 to 0.760 (mean r^2(cv) = 0.73; SD = 0.02) for five-component PLS models, suggesting only a minor dependence on the absolute orientation of the grid box. The effect of different choices of the origin of the grid was investigated using 14 atom types in addition to carbon C3 as probe atoms with a 2 Å grid spacing. The r^2(cv) values for each probe atom range from 0.743 to 0.769 for five-component PLS models with Cl or C3 and other carbon-based atom types showing the highest r^2(cv), showing only a slight dependence on the chosen probe (mean r^2(cv) = 0.75; SD = 0.01).

Biological activities were randomly assigned to molecular structures and subjected to a leave-one-out PLS analysis.[44] The mean r^2(cv) for 100 randomizations is

−0.05 (SD = 0.06; high 0.10, low −0.21), showing that the obtained model is significantly better than a random model. No model shows more than one significant PLS component.

Although cross-validation reflects the predictive power of a model, the leave-one-out method might produce too high r^2(cv) values. Thus, PLS analyses were run with two and five randomly chosen cross-validation groups containing randomly selected compounds for predicting. Because of the random group formation, both analyses were repeated 100 times. The mean r^2(cv) value of 0.738 for five PLS components and five cross-validation groups (SD = 0.02; high 0.788, low 0.682) is slightly lower than that from using the leave-one-out method with a low standard deviation, which also counts for a stable, predictive PLS model. When only two cross-validation groups were used, which corresponds to 69 randomly chosen compounds in the test and training sets each, a lower mean r^2(cv) value of 0.675 for five PLS components was observed (SD = 0.04; high 0.753, low 0.567), supporting the finding of a stable, significant, and predictive model for this dataset.

To estimate the influence of increasing information in *X* space on the predictive ability of the final CoMFA model, a validation strategy was used that is based on progressively dividing the full set into training and test datasets using statistical design procedures.[51] This stategy led to a rigorous evaluation of the model's external prediction capabilities. Representative training subsets containing between 20 and 136 members in increments of 4 compounds were chosen using a maximum dissimilarity algorithm[48,49] and 2D fingerprints, as described earlier.[51] For each designed training subset, a leave-one-out PLS analysis was used to extract the

Table 4. Fitted and Predicted Factor Xa Affinities from Various 3D-QSAR Models for 138 3-Amidinobenzyl-1H-indole-2-carboxamides and Derivatives[a]

compd	SAR_set	pK_i fXa	pred CoMFA A	pred CoMSIA C	pred CoMFA E	pred CoMSIA F	compd	SAR_set	pK_i fXa	pred CoMFA A	pred CoMSIA C	pred CoMFA E	pred CoMSIA F
1	1	2.77	2.98	3.18	2.55	2.92	71	1	1.88	2.02	2.28	2.05	1.99
2	1	2.30	2.60	2.65	2.50	2.53	72	1	1.45	1.40	1.71	1.39	1.48
3	2	3.07	3.33	3.08	2.35	2.32	73	1	2.99	2.85	2.73	2.73	2.80
4	2	3.89	3.72	3.56	2.52	2.82	74	1	2.85	3.08	2.78	2.87	2.86
5	2	3.30	3.92	4.04	4.01	4.10	75	1	3.12	2.70	2.74	2.74	2.82
6	2	3.59	3.84	3.91	3.80	3.94	76	1	3.05	3.29	3.31	3.30	3.27
7	1	2.77	2.65	2.55	2.75	2.68	77	1	1.75	1.76	1.87	1.60	1.67
8	2	2.52	2.07	2.41	2.06	2.22	78	1	3.21	2.74	3.06	3.07	3.20
9	2	4.05	3.83	3.86	3.79	3.93	79	1	3.09	2.74	2.74	3.07	3.19
10	2	3.40	3.36	3.55	3.16	3.32	80	1	3.18	3.14	3.11	3.16	3.11
11	1	2.46	2.62	2.40	2.64	2.55	81	1	2.85	3.33	3.07	3.33	3.08
12	1	2.80	3.05	3.31	3.10	3.02	82	1	3.55	3.30	3.44	3.27	3.40
13	1	2.22	1.91	1.70	2.15	2.09	83	1	3.85	3.82	3.72	3.84	3.66
14	1	2.44	2.46	2.14	2.42	2.34	84	1	2.84	3.05	3.21	3.12	2.91
15	1	2.82	2.53	2.33	2.51	2.49	85	1	3.85	3.46	3.54	3.67	3.86
16	1	3.00	2.83	2.61	2.84	2.71	86	1	2.65	2.87	2.83	2.95	2.78
17	1	2.96	2.84	2.53	2.83	2.52	87	1	4.15	3.65	3.48	3.67	3.57
18	1	3.04	3.00	3.23	2.98	3.11	88	1	3.03	3.01	3.06	2.96	3.11
19	1	1.32	1.38	0.84	1.51	1.19	89	1	3.22	3.26	3.35	3.28	3.37
20	1	1.43	1.44	1.41	1.40	1.53	90	1	3.08	3.29	3.52	3.44	3.52
21	2	3.60	3.36	3.37	2.97	3.19	92	1	3.62	3.31	3.41	3.46	3.41
22	2	3.30	3.51	3.35	3.16	2.85	93	1	2.87	3.24	3.07	3.29	3.14
23	2	3.02	3.11	3.25	2.78	3.07	94	1	2.40	2.36	2.71	2.41	2.62
24	1	3.17	2.93	2.86	2.72	2.88	95	1	2.59	2.82	2.84	2.77	2.70
25	2	0.84	1.26	1.47	2.13	2.48	96	1	2.88	3.18	2.73	3.08	3.05
26	2	2.04	2.03	1.83	2.13	2.03	97	1	3.05	3.03	2.72	2.86	2.92
27	1	2.15	2.08	1.91	2.27	2.06	98	1	2.68	2.90	2.50	2.85	2.83
28	1	1.79	1.94	1.84	1.94	1.71	99	2	3.35	3.45	3.38	3.38	3.14
29	1	1.81	2.00	1.86	1.95	1.91	100	1	1.44	1.58	1.36	1.41	1.06
30	1	2.01	1.77	1.49	1.85	1.75	101	1	2.74	2.58	2.26	2.23	2.32
31	2	1.61	2.05	2.14	1.30	1.66	102	2	1.79	2.04	2.55	2.62	2.97
32	1	1.89	1.97	2.55	1.88	2.18	103	2	1.87	2.09	2.25	2.30	2.94
33	2	2.50	2.06	2.10	1.28	1.65	104	1	1.27	1.50	1.42	1.48	1.61
34	2	3.00	2.90	2.74	2.23	2.46	105	2	2.56	2.67	2.87	2.28	2.83
35	1	2.08	2.36	2.18	2.11	2.22	106	2	2.49	2.56	2.59	2.48	2.56
36	2	1.92	2.09	2.44	1.87	2.85	107	1	1.16	1.42	1.49	1.33	1.44
37	2	1.76	1.86	2.11	2.17	2.21	108	2	1.69	2.12	2.51	2.21	2.79
38	2	2.08	1.57	1.82	2.20	1.95	109	2	1.95	2.17	2.39	2.21	2.73
39	2	0.76	0.81	1.16	1.62	2.19	110	1	1.25	1.00	1.02	1.06	1.17
40	2	3.00	3.47	3.26	2.49	2.72	111	2	1.10	1.11	1.32	1.52	1.89
41	2	3.60	3.10	3.04	2.58	2.57	112	2	1.15	0.98	1.11	1.31	1.82
42	2	2.10	1.75	1.82	2.05	2.29	113	2	3.82	4.09	3.49	3.05	2.58
43	2	2.72	2.48	2.51	2.27	2.65	114	2	3.70	3.48	3.35	3.01	3.25
44	2	2.71	3.11	2.61	2.34	2.82	115	1	1.62	1.81	1.65	1.48	1.57
45	2	4.05	3.71	3.48	2.74	3.00	116	1	1.17	1.35	1.38	1.41	1.42
46	2	3.96	3.65	3.27	2.56	2.93	117	1	1.15	1.09	1.29	1.03	1.20
47	1	4.15	3.68	3.70	3.67	3.63	118	2	1.15	1.40	1.31	1.89	1.91
48	1	3.21	3.31	3.23	3.46	3.50	119	1	1.69	2.14	2.10	1.94	1.93
49	1	3.29	3.39	3.65	3.36	3.45	120	2	1.48	2.01	2.04	2.19	2.60
50	1	4.05	3.82	3.60	3.77	3.49	121	2	2.52	2.70	2.44	2.67	2.76
51	1	3.11	3.24	3.59	3.34	3.42	122	2	2.76	2.79	2.43	2.31	1.93
52	1	2.88	2.71	2.84	2.76	3.11	123	1	1.06	1.28	1.15	1.18	1.14
53	1	3.72	3.76	3.93	3.55	3.89	124	1	1.12	1.27	1.04	1.22	1.16
55	2	2.93	2.50	2.65	2.32	2.31	125	1	3.55	3.60	3.41	3.53	3.29
56	2	2.74	2.48	2.55	2.50	3.67	126	1	3.36	3.61	3.64	3.49	3.43
57	1	2.62	2.54	2.41	2.40	2.63	127	1	0.26	0.52	0.39	0.54	0.41
58	2	2.58	2.39	2.57	2.28	2.46	128	1	1.38	1.12	1.20	1.17	1.35
59	2	3.41	2.88	3.11	2.45	2.81	129	1	2.02	1.95	1.63	2.05	1.79
60	2	2.14	2.13	2.43	2.20	2.64	130	1	0.89	0.78	0.72	0.76	0.79
61	2	1.00	1.32	1.41	1.33	2.39	131	1	1.00	1.05	1.09	1.17	1.30
62	2	3.38	2.67	2.66	2.27	1.97	132	2	1.52	1.31	1.09	1.71	1.68
63	2	2.51	2.52	2.50	2.40	3.01	133	1	1.41	1.35	1.17	1.16	1.08
64	2	1.69	1.45	1.68	1.82	2.68	134	1	2.21	2.17	1.91	1.99	1.82
65	2	3.89	3.60	3.43	2.51	2.85	135	1	1.88	1.98	2.30	2.14	2.17
66	2	2.88	2.62	2.86	2.44	2.88	136	1	0.85	0.48	0.84	0.54	0.74
67	2	1.00	1.26	1.04	1.94	3.84	137	1	1.66	1.63	1.80	1.72	1.78
68	2	2.32	2.62	3.04	2.41	2.80	138	1	1.74	1.62	2.02	2.00	1.95
69	1	2.61	3.16	2.92	2.99	2.94	139	1	1.91	1.61	1.96	1.92	1.89
70	1	3.64	3.68	3.43	3.61	3.48	140	2	1.95	1.70	1.88	1.92	1.70

[a] Experimental biological activity pK_i_fXa is expressed as $\log[(1/K_i)100000]$. See text and Table 3 for details on 3D-QSAR models. Predictions for all compounds from models A and C and for SAR set 1 for models E and F were based on the corresponding non-cross-validated PLS models with optimal number of components, while compounds indicated as SAR set 2 for models E and F were used as the external prediction set.

r^2(cv) for the optimal number of components. Then a PLS analysis without cross-validation led to the conventional r^2 and a model for predicting the remaining compounds. The prediction of all remaining compounds then results in the predictive r^2 value for each model. Those different r^2 values are plotted in Figure 4A versus the number of compounds for each designed training subset following this approach. The r^2(cv) value is low
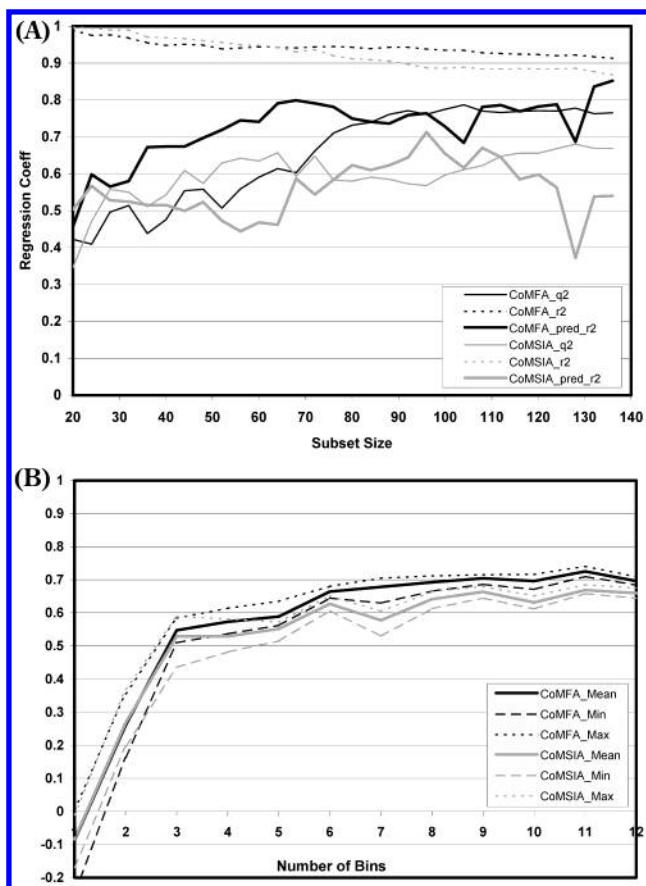
Figure 4. Validation and predictive ability assessment of factor Xa 3D-QSAR models. (A) Several smaller training sets designed using maximum dissimilarity methods. The graph shows a comparison of cross-validated, conventional, and predictive r^2 values for individual PLS models for a particular training set on the y axis versus the training set size using CoMFA or CoMSIA as the descriptor for PLS. (B) Progressive shuffling on y space to probe tolerance of the final PLS models against uncertainties in biological data. The number of bins for shuffling is plotted versus the mean, minimum, and maximum r^2(cv) values from repeated leave-one-out PLS analyses for both CoMFA and CoMSIA descriptors. For each individual bin, the results for 10 individual randomizations are averaged for analysis.

with less than 44 diverse compounds in the training set. Although the conventional r^2 is high, the predictive ability is not sufficient. The mean Tanimoto coefficient for this 44-compound training set is 0.83, while a pair of the most similar compounds in this subset show a Tanimoto coefficient of 0.88. When the subset size is increased, the cross-validated r^2 reaches values between 0.55 and 0.78. For the corresponding external predictions, the predictive r^2 is between 0.67 and 0.85, demonstrating stable models with good predictive capabilities for these external test sets. This example suggests that a reliable prediction of biological affinities can be expected for novel candidate molecules with a larger similarity than defined by this maximum Tanimoto threshold of ~0.88. The degree of extrapolation increases with decreasing similarity, thus leading to less reliable affinity predictions for external test sets.

Progressive shuffling[45,52] was applied to assess the stability of final models against variations in biological data, the Y space. After the activity is divided into 2−12 groups, each group is internally randomized while the

interrelationship remains unaffected. When 12 groups are used, only a small uncertainty is probed, while for 2 subgroups, a much larger portion of the y block is randomized (Figure 4B). This interpretation aids in understanding the effect of biological errors on PLS predictions and provides an estimate of model stability. With one group, the results are similar to complete randomization; negative r^2(cv) values are observed. The increase of the mean r^2(cv) with more subgroups is obvious, suggesting that shuffling using more than three subgroups still produces models of remarkable quality ($q^2 > 0.5$), thus suggesting that the models are less dependent on minor biological variations.

3.4. CoMSIA Model and Validation Studies. Similar statistical results were obtained when analyzing this dataset using CoMSIA steric, electrostatic, and hydrophobic fields. When a 2 Å grid spacing was used, a CoMSIA model with an r^2(cv) value of 0.655 for five PLS components and a conventional r^2 of 0.863 was obtained (model C, Table 3). The steric field descriptors (2156 variables) explain only 26% of the variance, while the proportion of the electrostatic descriptors remains the same at 43%. Now the additional hydrophobic field explains the remaining 31% of the variance. Thus, the CoMFA steric field contribution can be seen as a balance between pure steric plus hydrophobic effects. With a grid spacing of 1 Å (16 632 variables, identical region definition compared to CoMFA 1 Å), the model remains stable; a five-component model was obtained with r^2(cv) of 0.657 and r^2 of 0.866 (model D, Tables 3 and 4).

The CoMSIA 2 Å model was also subjected to further validation studies. For CoMSIA, the effect of the alignment relative to the grid position and box was evaluated by consistently moving all compounds in increments of 0.5 Å, although this did not affect the PLS models, as confirmed by r^2(cv) values of 0.655 (five components) for each orientation. This might be explained by the Gaussian-type smoothing function to compute CoMSIA fields.[53]

When the biological activities are randomized 100 times, a mean r^2(cv) of −0.07 (SD = 0.07; high 0.08, low −0.32) is observed, revealing the significance of this model. When over 100 PLS analyses are averaged with two random cross-validation groups, a mean r^2(cv) of 0.590 (SD = 0.04; high 0.695, low 0.498) results for the five-component models. This value is even increased when using five cross-validation groups. Here, a mean r^2(cv) value of 0.635 for five PLS components and five cross-validation groups results (SD = 0.03; high 0.691, low 0.537), which again is only slightly lower than the r^2(cv) value obtained using the leave-one-out procedure.

For evaluation of the final CoMSIA model predictivity, the same maximum-dissimilarity-selected training and test sets compared with the CoMFA validation study described above were used. Again, the obtained r^2 values for each subset are plotted in Figure 4A on the y axis versus the number of compounds in the diverse training sets. For CoMSIA, the first reliable PLS models emerge with 64 training set molecules, corresponding to a Tanimoto similarity threshold of 0.92 for reliable prediction and a mean Tanimoto coefficient of 0.88 for this particular training subset. In general, both the cross-validated and predictive r^2 values are lower for CoMSIA, although there are no large differences for training sets with more than 100 compounds. The predictive r^2 value

for external predictions based on designed training subsets is between 0.5 and 0.7. The predictive r^2 value for the CoMSIA models decreases again with more than 132 molecules, as here the test set is statistically no longer significant.

This rigorous validation study based on several external test sets suggests that the CoMFA models have a higher predictive power than CoMSIA for this particular case, while CoMSIA is less dependent on alignment inaccuracies and thus might be comparable from a practical point of view for designing novel candidate molecules. For each training set with acceptable statistical parameters after PLS analysis, we checked whether its chemical interpretation in terms of CoMFA and CoMSIA contour maps led to similar conclusions as reported for the entire 3-amidinobenzyl-1*H*-indole-2-carboxamide dataset. This was always the case. We consider this to be the most relevant information that could be extracted from these 3D-QSAR models in conjunction with reliable affinity prediction for novel candidates.

Progressive shuffling[52] applied to the final CoMSIA model also demonstrates its robustness against variations in biological data (Figure 4B). The increase of the mean r^2(cv) values with more than one subgroup is obvious, suggesting that shuffling using more than three subgroups still produces models of remarkable quality for CoMSIA ($q^2 > 0.5$). This analysis shows that both QSAR techniques CoMFA and CoMSIA are robust and can tolerate minor biological variations.

3.5. Prospective Design Based on 3D-QSAR Models. A final 3D-QSAR model validation was done using a prospective study with an external test set. When only 82 compounds from this dataset were used, which were synthesized during the initial phase of our fXa lead optimization project (SAR_set 1 in Table 4), 3D-QSAR models based on CoMFA and CoMSIA were derived and were statistically validated in a manner similar to that described above and employed for prioritization of novel synthesis candidates. One example is given in ref 22, where this model in combination with flexible docking was used to select appropriate S4-directed building blocks from the ACD database.[54] By use of solid-phase chemistry,[22] a variety of novel hydrophobic S4 substituents were added to the 3-amidinobenzyl-1*H*-indole-2-carboxamide scaffold (Schemes 2 and 3).

When a 2 Å grid spacing was used, a CoMFA model with an r^2(cv) value of 0.698 for four relevant PLS components and a conventional r^2 of 0.938 were obtained for those 82 compounds (model E, Tables 3 and 4). Here, steric field descriptors contribute 54% to the total variance, while the electrostatic field explains 46%. The corresponding CoMSIA model with 2 Å grid spacing led to an r^2(cv) value of 0.660 for five PLS components and a conventional r^2 of 0.933 (model F, Tables 3 and 4). The contributions for steric, electrostatic, and hydrophobic fields are 25%, 44%, and 31%, respectively. The chemical interpretations of both the CoMFA and CoMSIA models were similar to the interpretation given in the following section for the entire dataset. However, at this time of our fXa project, there was a need to replace basic S4-directed substituents against more hydrophobic building blocks in order to improve pharmacokinetic properties. Hence, the external prediction

test set, consisting of 57 compounds, shows somewhat different properties; this also causes external predictions to be less accurate than those based on statistically designed training sets.

Prospective predictions for 56 synthesized compounds in Table 4 resulted in a predictive r^2 value of 0.502 for CoMFA model E with 2 Å grid spacing, which is acceptable given the property change in S4-directed substituents. The corresponding predictions for 138 compounds from PLS models A and C as well as for submodels E and F are summarized in Table 4. All compounds marked as SAR set 2 in this table were used for external prediction. The corresponding predictive r^2 value of 0.162 for CoMSIA is much lower than that observed for CoMFA. This is in line with the designed training subset studies, showing CoMSIA to have lower predictive capabilities in this case. This lack of predictivity here has to be attributed to our attempt to change physicochemical properties in the S4-directed region. However, the quality of external predictions was acceptable for CoMFA regardless of this property change. Indeed, this model was used for subsequent design cycles to synthesize and profile novel candidate molecules. Here, a combination of structure-based design with simultaneous inspection of 3D-QSAR contours was effective to support the design process to search for novel inhibitors. Many novel synthesis proposals were generated, evaluated using the outlined procedure and reliably ranked by 3D-QSAR affinity predictions. Each design cycle was then completed by generation and application of novel, improved 3D-QSAR models. This stepwise design and synthesis procedure based on CoMFA predictions constantly improved statistical results and predictivity in this series.

3.6. Comparison with the fXa Binding Site. The steric and electrostatic std*coeff fields for the final CoMFA analysis A of 138 molecules with 2 Å grid spacing based on the factor Xa binding site topology are displayed in Figure 5 with the inhibitor **41** (K_i(fXa) = 25 nM) carrying a hydrophobic 3,5-dichlorobenzyl substituent in the S4 pocket. In Figure 5A, steric field contributions correlated with changes in biological activity are displayed. Green contours (>85% contribution) indicate regions where steric bulk is favorable for new inhibitors, while yellow contours (<15% contribution) highlight regions where bulky substituents are detrimental for biological activity. A similar analysis is given for the electrostatic SD*coeff field (Figure 5B). Blue contours (>85% contribution) refer to regions where an increase of positive charge (or a decrease of negative charge) is favored for new ligands to enhance affinity, while red contours (<15% contribution) indicate those regions where an increase of negative charge is favorable for biological activity. Those 3D-QSAR results were derived by taking only ligand information from the 3-amidinobenzyl-1*H*-indole-2-carboxamide series into account, which were generated by docking and minimization of protein−ligand complexes. However, the resulting PLS contour maps are discussed in combination with the fXa topology to underscore consistency with steric, electrostatic, and hydrophobic requirements imposed from this binding site. To support this combined interpretation, the steric and electrostatic CoMFA contour maps are additionally mapped onto the solvent-
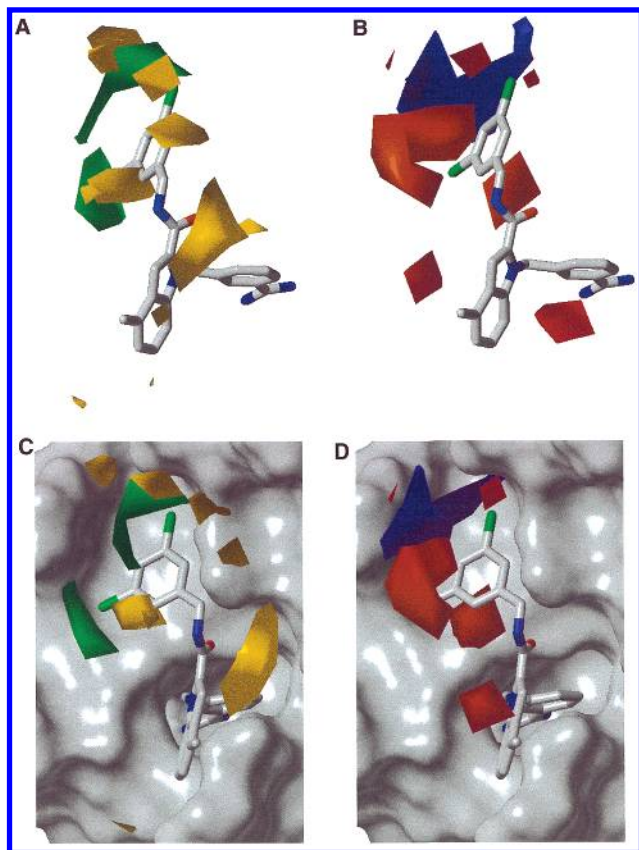
Figure 5. Contour maps from the final CoMFA analysis A with 2 Å grid spacing in combination with the inhibitor **41**, where hydrogens are omitted for clarity. (A) Steric std*coeff contour map. Green contours (>85% contribution) refer to sterically favored regions, and yellow contours (<15% contribution) indicate unfavored areas. (B) Electrostatic std*coeff contour map. Blue contours (>85% contribution) refer to regions where negatively charged substituents are unfavorable, and red contours (<15% contribution) indicate regions where negatively charged substituents are favorable. (C) Same as in (A) with factor Xa solvent-accessible surface showing complementarity to receptor topology. (D) Same as in (B) with factor Xa solvent-accessible surface.

accessible surface of the fXa binding site in parts C and D of Figure 5.

Steric bulk is favorable at positions 3 and 5 of the distal ring for the 3-amidinobenzyl-1*H*-indole-2-carboxamide series situated in the fXa S4 pocket, which is surrounded by three aromatic residues (Trp215, Phe174, and Tyr99). Interestingly, purely hydrophobic substitutions at both meta positions of this ring also led to highly active compounds, e.g., dichloro or dimethyl groups as in compounds **41** and **45** with K_i values of 25 and 9 nM, respectively. Interestingly, there is one example reported in the literature with a 2,6-dimethylpiperidyl moiety in the S4 aryl binding site.[16m] The relatively good biological activity of inhibitors around this scaffold with purely hydrophobic S4-directed substituent is in contrast to earlier findings, where only basic or cationic substituents and/or hydrogen-bond donors are reported to lead to high activity in this S4 aryl binding subsite. High affinity in the 3-amidinobenzyl-1*H*-indole-2-carboxamide series is observed for substituents such as 3-amidinobenzyl (77 nM, **51**) and 4-(trimethylammonium)benzyl (7 nM, **87**). A 3-chlorobenzyl substituent with only a single chlorine atom in S4 shows a K_i value of

262 nM (**58**), while the corresponding halogen at the 2-position of the benzyl ring is much less active (1200 nM, **36**). Changing chlorine against iodine improves affinity (42 nM, **62**), while adding the second chloro substituent in the 5-position results in a compound with a K_i value of 25 nM (**41**, CH$_3$ instead of OH at position 4). For comparison, the unsubstituted benzyl group alone is much less active (2032 nM, **108**). Substituted benzyl moieties provide optimal geometry to orient those hydrophobic substituents within the S4 pocket, while shorter substituents such as 3,5-dichlorophenyl or longer residues such as corresponding phenethyl substituents are much less active (e.g., 798 nM, **42**) for this series.

One favorable steric contour is located at the more solvent-exposed region between the aromatic ring of Phe174 and the side chain of Gln217. This position accommodates the second hydrophobic methyl or chlorine substituents in the fXa−**41** and fXa−**45** complexes, as discussed above and thus is essential for high binding affinity for 3-amidinobenzyl-1*H*-indole-2-carboxamides. The other contour region points toward the edge of S4 close to residues Glu97−Thr98, where an important structurally conserved water is present even in structures with more hydrophobic substituents. This position accommodates the other hydrophobic chlorine or methyl substituent in the X-ray structures of compounds **41** or **45**. In the rear of the S4 pocket, several backbone amide carbonyl groups are oriented toward the interior, forming a distinct hydrogen bonding acceptor pocket pointing toward the structurally conserved water molecule.

In contrast, the addition of steric bulk directed toward Tyr99, separating the S4 from the smaller S2 subsite, is unfavorable for biological activity because this would shift the aromatic ring of Tyr99 toward His57 of the protease catalytic triad. For example, while the 1-naphthyl substituent at the 3-amidinobenzyl-1*H*-indole-2-carboxamide scaffold nicely fits into the S4 pocket (68 nM, **24**), the corresponding 2-naphthyl substituent is in steric contact with Tyr99, leading to a less active inhibitor (1720 nM, **37**). Other examples of bulky substituents (also with unfavorable electronic properties) are **103**, with a 4-methoxybenzyl group (1360 nM), and **39**, with a 3,4-dimethoxybenzyl substituent (17 350 nM), while the corresponding 3,4-dichlorobenzyl residue even with the unfavorable 3,4-orientation of both halogens shows a higher activity (841 nM, **38**). There are other examples of either larger substituents attached to the distal ring or a longer spacer between the 3-amidinobenzyl-1*H*-indole-2-carboxamide and the distal aromatic ring (e.g., C$_2$H$_4$ instead of CH$_2$), shifting otherwise identical substituents toward the rear of the fXa S4 pocket, which causes the affinity to decrease. This is exemplified by comparing **58** with a 3-chlorobenzyl residue (262 nM) versus **27** with a 2-(3-chlorophenyl)ethyl) substituent (707 nM).

Interestingly, the protein topology of the aryl binding pocket for this ligand series consistently differs from protein structures experimentally observed for other scaffolds[19] and is more similar to 1HCG,[18] where fXa is complexed with a second fXa molecule. In other fXa structures, this S4 subsite is much smaller because of a movement of the Tyr99 aromatic ring toward Phe174 at the other edge of this subsite.[19] However, even the herein-observed enlarged subsite topology led to active
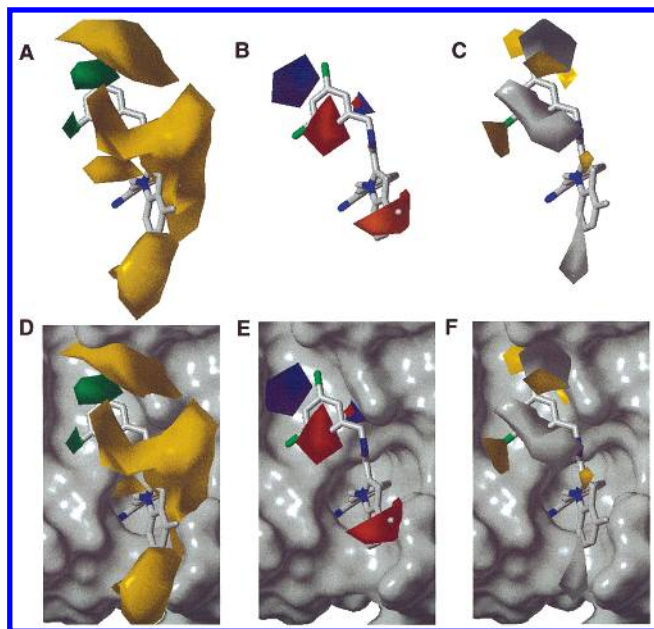
Figure 6. Contour maps from the final CoMSIA analysis C with 2 Å grid spacing in combination with the inhibitor **41**, where hydrogens are omitted for clarity. (A) Steric std*coeff contour map. Green contours (>85% contribution) refer to sterically favored regions, and yellow contours (<15% contribution) indicate unfavored areas. (B) Electrostatic std*coeff contour map. Blue contours (>85% contribution) refer to regions where negatively charged substituents are unfavorable, and red contours (<15% contribution) indicate regions where negatively charged substituents are favorable. (C) Hydrophobic std*coeff contour map. White contours (>80% contribution) refer to regions where hydrophilic substituents are favorable, and yellow contours (<20% contribution) indicate regions where hydrophobic substituents are favorable. (D) Same as in (A) with factor Xa solvent-accessible surface to illustrate complementarity to receptor topology. (E) Same as in (B) with factor Xa solvent-accessible surface. (F) Same as in (C) with factor Xa solvent-accessible surface.

factor Xa inhibitors carrying only hydrophobic substituents directed toward S4. Other sterically unfavorable regions are located at both ortho positions of the distal ring in S4. When another chlorine at this position is added to compound **41** (25 nM), the fXa activity is reduced to 99 nM for the 2,3,5-trichlorobenzyl substituent (**34**). The steric contour maps also indicate that modifications or substitutions at the CH_2 linker connecting the amide nitrogen to the distal phenyl ring are unfavorable for activity because of the Tyr99 side chain limiting this subsite. Another sterically unfavorable region is the upper site of indole position 3 directed toward the active site and the flexible side chain of Gln192. Small, hydrophobic substitutions at this position are favorable, like bromine in **4** (13 nM), while larger groups did not lead to improved binding affinity. This is exemplified by the series of indole-3-carboxamides (**137**−**140**) with activities greater than 1100 nM.

The chemical interpretation of the CoMSIA model C with 2 Å grid spacing is highly consistent with that of CoMFA and also in agreement with the experimental factor Xa binding topology. The corresponding contour maps for steric fields are displayed in Figure 6 without (Figure 6A−C) and with (Figure 6D−F) a binding site solvent-accessible surface. Again, green contours from the PLS derived std*coeff field (>85% contribution)

indicate those regions where steric bulk is favorable, while yellow contours (<15% contribution) highlight regions where steric bulk is detrimental to activity.

In addition to CoMFA, a sterically unfavorable region close to the indole position 5 is present, which relates to 3-amidinobenzyl-1*H*-indole-2-carboxamide derivatives with bulky substituents such as *O*-benzyl at this position, as **52** (133 nM) compared to the unsubstituted *m*-bisamidine **53** (19 nM as the di-TFA salt and 26 nM as diactetate **54**). A corresponding CoMFA contour is present at lower contour levels. This reduction in affinity is not observed for the corresponding 4-(trimethylammonium)benzyl substituent directed toward S4. Here, the *O*-benzyl derivative **79** and the unsubstituted **76** have similar affinities (82 versus 90 nM), respectively, which might be due to the nonoptimal binding geometry of para-substituted S4-directed substituents for this scaffold series. This lower affinity for favorable S4 substituents might be attributed to unfavorable steric interactions of the bulky *O*-benzyl substituent with the Arg143 side chain, as can be deduced in the fXa−**79** X-ray structure (see above) and docking binding modes for the corresponding analogues. A detailed inspection of the experimental protein structure shows that the Arg143 side chain is shifted because of this substitution by a modification of the $\chi3$ dihedral angle. This leads to the formation of a hydrogen bond between the side chains of Arg143 and Gln192, while the conformation of the latter residue also has to be shifted, which might be less favorable for binding affinity.

The inspection of the CoMFA and CoMSIA electrostatic contour maps in Figure 5B,D and Figure 6B,E shows that more positive charge at the rear of the S4 subsite is favorable for biological activity in the 3-amidinobenzyl-1*H*-indole-2-carboxamide series. This finding is in agreement with the high affinities for bisamidines and other S4-directed substituents with basic or cationic character. The *m*-bisamidine **53** and the 4-(trimethylammonium)benzyl-substituted **76** have acceptable K_i values of 19 and 90 nM, respectively. In contrast, the corresponding S4-directed unsubstituted benzyl group in **107** led to a much reduced affinity (6890 nM). This loss of favorable electrostatic interaction energy has to be compensated by sterically adequate substitutions. Many X-ray structures and their analysis further underscore the favorable nature of cation−π interactions[18,19,21,55−58] between the ligand substituent and the Trp215 indole side chain. For this S4 subsite, it has been reported that quaternary amines such as the trimethylphenylammonium cation are able to inhibit factor Xa with K_i values in the low millimolar range.[56] This weak inhibitor and the S1/S4 directed 3-amidinobenzyl-1*H*-indole-2-carboxamides presented here with higher affinities thus exhibit favorable hydrophobic and van der Waals interactions with the aromatic side chains of Phe174, Trp215, and Tyr99 surrounding the S4 pocket, and the charged amine or amidine residues are in favorable electrostatic interactions with the highly flexible side chain of Glu97 at the edge of S4, which is able to fold over the S4 pocket. Thus, the detailed analysis of experimental binding modes has revealed three subregions as key ligand binding domains within S4: the "hydrophobic box" formed by the known three aromatic amino acids forming a deep aryl binding

pocket, the "cationic hole" close to Glu97, and the water binding site at the rear of S4.

In contrast, red contours for CoMFA and CoMSIA point to those regions in the S4 and S3 pocket where an increase in negative charge causes favorable interactions with the protein site. In particular, more negative charge on both sides of the distal aromatic substituent and more negative charge at the ortho position of this ring oriented toward the S3 subsite close to the Gln217 side chain and Gly216 are favorable in the 3-amidinobenzyl-1*H*-indole-2-carboxamide series. Interestingly, the interpretation of CoMSIA contours shows another favorable region where more negative or less positive charge is favorable for high affinity located at positions 4 and 5 of the indole ring. In this series, this is exemplified by **49** with an amino group attached to position 5 compared to the unsubstituted bisamidine **53** with 19 nM affinity. The corresponding OH group at this position led to a K_i value of 77 nM in compound **51**, while shifting the substituent to position 4 led to a higher affinity in this series (e.g., methyl in **47** with 7 nM, OH in **46** with 11 nM). In the protein structure, this region is located close to Arg143 and Gln192, suggesting unfavorable interactions with both side chain functionalities with more positively charged substituents at the indole scaffold in this series. A corresponding CoMFA contour is present at a lower contour level. This CoMSIA interpretation adds additional information on the unfavorable mobility of Arg143 and Gln192 caused by bulky substituents at the indole ring and details about favorable and unfavorable hydrophobic field distributions for 3-amidinobenzyl-1*H*-indole-2-carboxamides.

To estimate entropic contributions to ligand binding in factor Xa, additional hydrophobic CoMSIA fields were computed and used for PLS analysis. White contour regions in Figure 6C,F indicate regions where hydrophilic interactions are favorable for biological activity (>85% contribution). Those regions are located at the rear of the S4 subsite, where protein carbonyl groups point inward to form hydrogen-bonding networks, and at the ortho position of the distal ring of the 3-amidinobenzyl-1*H*-indole-2-carboxamide series located in S4 directed toward the backbone of Gly216 and Gln217.

Preferred hydrophobic interactions are indicated by yellow contour regions in Figure 6 (<15% contribution). They are located within the S4 pocket and, in addition, close to position 3 of the indole scaffold in this series, where methyl or halogen substitutions are known to have a favorable effect on factor Xa biological affinity compared to unsubstituted 3-amidinobenzyl-1*H*-indole-2-carboxamide derivatives. This is exemplified by compound **4** with a bromine at C3 and 13 nM affinity. Furthermore, the affinity of the unsubstituted *m*-bisamide **53** (19 nM) was improved by adding a chlorine in this position (9 nM, **9**), while bromine did not lead to higher affinity. Hence, the combined interpretation of the derived CoMFA and CoMSIA contour maps is in good agreement with the protein binding topology and the experimentally determined X-ray structures of protein–ligand complexes for the 3-amidinobenzyl-1*H*-indole-2-carboxamide scaffold.

This detailed analysis of molecular interaction fields shows that it is possible to replace cationic and polar functionalities by adequate hydrophobic substituents in the S4 pocket caused by different interaction subsites within this protein structural motif.

3.7. Selectivity toward Thrombin and Trypsin. A set of 24 3-amidinobenzyl-1*H*-indole-2-carboxamide based factor Xa inhibitors with acceptable biological affinity was tested against the related protease thrombin (Table 1). All compounds except **65**, **113**, **114**, **46**, and **34** were selective with K_i values greater than 1 μM for thrombin. Compound **65** with a 3-methyl-5-(trifluoromethyl)benzyl substituent directed toward S4 and an OH group attached to the indole position 4 shows a K_i value of 480 nM against thrombin and 13 nM against fXa. Changing the CF$_3$ substituent against CH$_3$ in this series slighly increases the fXa affinity (**46**, 11 nM), while the thrombin affinity decreases (620 nM). Interestingly the substitution of hydroxy versus methyl at indole position 4 at the 3-amidinobenzyl-1*H*-indole-2-carboxamide scaffold produces more selective fXa inhibitors (**45**, 9 vs 1420 nM for dimethyl substituents in S4), while a 3,5-di(trifluoromethyl)benzyl substitution causes less selective inhibitors (**40**, 99 vs 1180 nM). The isoquinoline-1-yl-methyl substituent present in molecules **113** and **114** also does not lead to selective inhibitors with fXa affinities of 15 and 20 nM or thrombin affinities of 200 and 290 nM, respectively. The least selective S4-directed substituent in the 3-amidinobenzyl-1*H*-indole-2-carboxamide series is the 2,3,5-trichlorobenzyl building block in **34** with 99 nM for Xa and 270 nM for thrombin. In agreement with a comparison of fXa and thrombin active site topologies,[59,60] more selective fXa inhibitors are obtained with basic substituents situated in the S4 pocket. A quantitative analysis of thrombin affinity and thrombin–factor Xa selectivity will be given elsewhere.[61] Hence, several compounds from this 3-amidinobenzyl-1*H*-indole-2-carboxamide series are potent and selective toward the related protease thrombin.

Trypsin is another serine protease related to factor Xa and thrombin. Thus, the X-ray structure of **79** was determined in trypsin as a surrogate for factor Xa before factor Xa crystals were available. As shown in Figure 2C, the binding modes of **79** in factor Xa and trypsin were similar. The major difference is the absence of the *O*-benzyl group at the indole ring in trypsin, which was not visible in the electron density maps most likely because of disorder. In trypsin, the terminal trimethylaniline group is situated deeper in the more open S4 pocket. Here, Tyr99 in factor Xa is replaced with Ile, and Phe174 is replaced with Gly in trypsin, while the loop 172–175 is rotated away from the trypsin S4 pocket, making this subsite larger.

By inspection of both protein–ligand complexes, several factors contributing to the weaker binding affinity of **79** in trypsin (1507 nM compared to 82 nM in fXa) were identified. First, the **79** *O*-benzyl group is disordered in trypsin, which does not support strong binding. In contrast, the *O*-benzyl group shows favorable π interactions with the guanidine group of Arg143 in factor Xa, which is replaced with Asn in the corresponding trypsin structure. While the inhibitor–enzyme interactions are similar for the benzamidine–indole part of the inhibitor, there are further differences in the S4 pocket. In factor Xa, the trimethylaniline group is bound

in a hydrophobic pocket between Tyr99 and Phe174. In trypsin, the S4 pocket is more solvent-exposed with only Trp215 conserved at its bottom. Furthermore, it is known that the fXa S4 pocket is more negatively charged than the S4 pocket of trypsin,[19,21,55] which would bind the positively charged trimethylamine group more strongly.[56] This is in perfect agreement with biological affinities for **79** with both enzymes. A more quantitative analysis of binding site differences for both proteases will be given elsewhere.[61]

4. Conclusions

Although structure-based design approaches are focused toward a detailed understanding of relevant protein−ligand interactions, those do not always allow us to derive predictive models. Complementary statistical 3D-QSAR approaches, on the other hand, are often limited; they cannot be interpreted in protein structural terms because the derived statistical models are not always based on the "bioactive conformation" of a particular scaffold. In this study, we successfully combined both approaches for a detailed understanding of ligand and protein structural requirements for potent factor Xa inhibitors in the 3-amidinobenzyl-1*H*-indole-2-carboxamide series. On the basis of several factor Xa X-ray structures for 3-amidinobenzyl-1*H*-indole-2-carboxamides, consistent and highly predictive 3D-QSAR models were derived, which could be analyzed and chemically interpreted in terms of important protein−ligand interactions for this scaffold. The resulting models provide guidelines for further ligand design in this 3-amidinobenzyl-1*H*-indole-2-carboxamide series plus a predictive score for synthetic candidates within the prediction frame of the 3D-QSAR technique. Of course, 3D-QSAR techniques only allow for interpolation, while some of the herein presented validation techniques allow us to estimate the similarity threshold for novel compounds, allowing for a more reliable affinity prediction.[62] The structural and chemical interpretation of CoMFA and CoMSIA contour maps directly points to those regions in the factor Xa binding site where steric, electronic, or hydrophobic effects play a dominant role in ligand−receptor interactions. Extensive statistical testing of the models has validated their robustness, while a prospective design study was based on an earlier 3D-QSAR model to prioritize novel substituents with altered physicochemical properties in the S4 pocket. The combined interpretation of 3D-QSAR derived contour regions for relevant protein−ligand interaction features being favorable for affinity in combination with the information from X-ray structural data for this series allows us to unravel the main features for a successful design of potent factor Xa inhibitors around the presented series. The comparable results from 3D-QSAR approaches and their complementarity to the receptor topology further indicate the validity of these models. From these analyses, a series of potent, achiral, and selective factor Xa inhibitors emerged, and this offers opportunities for compound optimization.

References

(1) Vlasuk, G. P. Direct Factor Xa inhibitors. In *New Therapeutic Agents in Thrombosis and Thrombolysis*; Sasahara, A. A., Loscalzo, J., Eds.; Marcel Dekker: New York, 1997; pp 261−283.

(2) *Hemostasis and Thrombosis: Basic Principles and Clinical Practice*; Coleman, R. W., Marder, V. J., Salzman, E. W., Eds; J. B. Lippincott: Philadelphia, PA, 1994.

(3) Davie, E. W.; Fujikawa, K.; Kisiel, W. The coagulation cascade: initiation, maintenance, and regulation. *Biochemistry* **1991**, *30*, 10363−10370.

(4) Mann, K. G.; Nesheim, M. E.; Church, W. R.; Haley, P.; Krishnaswamy, S. Surface dependent reactions of the vitamin K-dependent enzyme complexes. *Blood* **1990**, *76*, 1−16.

(5) Ahmad, S. S.; Rawala-Sheikh, R.; Walsh, P. N. Compounds and assembly of the factor Xa activating complex. *Semin. Thromb. Hamostasis* **1992**, *18*, 311−323.

(6) Harker, L. A.; Hanson, S. R.; Kelly, A. B. Antithrombotic Strategies Targeting Thrombn Activities, Thrombin Receptors and Thrombin Generation. *Thromb. Haemostasis* **1997**, *78*, 736−741.

(7) Waxman, L.; Smith, D. E.; Arcuri, K. E.; Vlasuk, G. P. Tick anticoagulant peptide (TAP) is a novel inhibitor of blood coagulation factor Xa. *Science* **1990**, *248*, 593−596.

(8) Hara, T.; Yokoyoma, A.; Tanabe, K.; Ishihara, H.; Iwamoto, M. DX-9065a, an orally active, specific inhibitor of factor Xa inhibits thrombosis without affecting bleeding time in rats. *Thromb. Haemostasis* **1995**, *74*, 635−639.

(9) Antman, E. M. Hirudin in acute myocardial infarction. Safety report from the thrombolysis and thrombin inhibition in myocardial infarction (TIMI) 9A trial. *Circulation* **1994**, *90*, 1624−1630.

(10) Philippides, G. J.; Loscalzo, J. Potential advantages of direct-acting thrombin inhibitors. *Coron. Artery Dis.* **1996**, *7*, 497−507.

(11) Al-Obeidi, F.; Ostrem, J. A. Factor Xa inhibitors. *Expert Opin. Ther. Pat.* **1999**, *74*, 635−639.

(12) Ewing, W. R.; Pauls, H. W.; Spada, A. P. Progress in the design of inhibitors of coagulation factor Xa. *Drugs Future* **1999**, *24*, 771−787.

(13) Zhu, B.-Y.; Scarborough, R. M. Factor Xa Inhibitors: Recent Advances in Anticoagulant Agents. *Annu. Rep. Med. Chem.* **2000**, *35*, 83−102.

(14) Fevig, J. M.; Wexler, R. R. Thrombin and Factor Xa Inhibitors. *Annu. Rep. Med. Chem.* **1999**, *34*, 81−100.

(15) Rai, R.; Sprengler, P. A.; Elrod, K. C.; Young, W. B. Perspectives on Factor Xa Inhibition. *Curr. Med. Chem.* **2001**, *8*, 101−119.

(16) (a) Pinto, D. J. P.; Orwat, M. J.; Wang, S.; Fevig, J. M.; Quan, M. L.; Amparo, E.; Cacciola, J.; Rossi, K. A.; Alexander, R. S.; Smallwood, A. M.; Luettgen, J. M.; Liang, L.; Aungst, B. J.; Wright, M. R.; Knabb, R. M.; Wong, P. C.; Wexler, R. R.; Lam, P. Y. S. Discovery of 1-[3-(Aminomethyl)phenyl]-*N*-[3-fluoro-2′-(methylsulfonyl)-[1,1′-biphenyl]-4-yl]-3-(trifluoromethyl)-1*H*-pyrazole-5-carboxamide (DPC423), a Highly Potent, Selective, and Orally Bioavailable Inhibitor of Blood Coagulation Factor Xa. *J. Med. Chem.* **2001**, *44*, 566−578. (b) Han, Q.; Dominguez, C.; Stouten, P. F. W.; Park, J. M.; Duffy, D. E.; Galemmo, R. A.; Rossi, K. A.; Alexander, R. S.; Smallwood, A. M.; Wong, P. C.; Wright, M. M.; Leuttgen, J. M.; Knabb, R. M.; Wexler, R. R. Design, Synthesis, and Biological Evaluation of Potent and Selective Amidino Bicyclic Factor Xa Inhibitors. *J. Med. Chem.* **2000**, *43*, 4398−4415. (c) Guilford, W. J.; Shaw, K. J.; Dallas, J. L.; Koovakkat, S.; Lee, W.; Liang, A.; Light, D. R.; McCarrick, M. A.; Whitlow, M.; Ye, B.; Morrissey, M. M. Synthesis, Characterization and Structure−Activity Relationships of Amidine-Substituted (Bis)benzylidene−Cycloketone Olefine Isomers as Potent and Selective Factor Xa Inhibitors. *J. Med. Chem.* **1999**, *42*, 5415−5425. (d) Quan, M. L.; Liauw, A. Y.; Ellis, C. D.; Pruitt, J. R.; Carini, D. J.; Bostrom, L. L.; Huang, P. P.; Harrison, K.; Knabb, R. M.; Thoolen, M. J.; Wong, P. C.; Wexler, R. R. Design and Synthesis of Isoxazoline Derivatives as Factor Xa Inhibitors. 1. *J. Med. Chem.* **1999**, *42*, 2752−2759. (e) Quan, M. L.; Ellis, C. D.; Liauw, A. Y.; Alexander, R. S.; Knabb, R. M.; Lam, G. N.; Wright, M. R.; Wong, P. C.; Wexler, R. R. Design and Synthesis of Isoxazoline Derivatives as Factor Xa Inhibitors. 2. *J. Med. Chem.* **1999**, *42*, 2760−2773. (f) Ewing, W. R.; Becker, M. R.; Manetta, V. E.; Davis, R. S.; Pauls, H. W.; Mason, H.; Choi-Sledeski, Y. M.; Green, D.; Cha, D.; Spada, A. P.; Cheney, D. L.; Mason, J. S.; Maignan, S.; Guilloteau, J.-P.; Brown, K.; Colussi, D.; Bentley, R.; Bostwick, J.; Kasiewski, C. J.; Morgan, S. R.; Leadley, R. J.; Dunwiddie, C. T.; Perrone, M. H.; Chu, V. Design and Structure−Activity Relationships of Potent and Selective Inhibitors of Blood Coagulation Factor Xa. *J. Med.*

Chem. **1999**, *42*, 3557–3571. (g) Choi-Sledeski, Y. M.; McGarry, D. G.; Green, D. M.; Mason, H. J.; Becker, M. R.; Davis, R. S.; Ewing, W. R.; Dankulich, W. P.; Manetta, V. E.; Morris, R. L.; Spada, A. P.; Cheney, D. L.; Brown, K. D.; Colussi, D. J.; Chu, V.; Heran, C. L.; Morgan, S. R.; Bentley, R. G.; Leadley, R. J.; Maignan, S.; Guilloteau, J.-P.; Dunwiddie, C.; Pauls, H. W. Sulfonamidopyrrolidinone Factor Xa Inhibitors: Potency and Selectivity Enhancements via P-1 and P-4 Optimization. *J. Med. Chem.* **1999**, *42*, 3572–3587. (h) Kucznierz, R.; Grams, F.; Leinert, H.; Marzenell, K.; Engh, R. A.; von der Saal, W. Tetrahydro-isoquinoline-based factor Xa inhibitors. *J. Med. Chem.* **1998**, *41*, 4983–4994. (i) Renatus, M.; Bode, W.; Huber, R.; Stürzebecher, J.; Stubbs, M. T. Structural and Functional Analyses of Benzamidine-Based Inhibitors in Complex with Trypsin: Implications for the Inhibition of Factor Xa, tPA, and Urokinase. *J. Med. Chem.* **1998**, *41*, 5445–5456. (j) Marlowe, C. K.; Uma, S.; Gunn, A. C.; Scarborough, R. M. Design, synthesis and structure–activity relationship of a series of arginine aldehyde factor Xa inhibitors. Part 1: Structures based on the (D)-Arg-Gly-Arg tripeptide sequence. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 13–16. (k) Gong, Y.; Pauls, H. W.; Spada, A. P.; Czekaj, M.; Liang, G.; Chu, V.; Colussi, D. J.; Brown, K. D.; Gao, J. Amido-(Propyl and Allyl)-hydroxybenzamidines: Development of Achiral Inhibitors of Factor Xa. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 217–221. (l) He, W.; Hanney, B.; Myers, M. R.; Spada, A. R.; Brown, K.; Colussi, D.; Chu, V. Nonbenzamidine compounds as selective factor Xa inhibitors. *Bioorg. Med. Chem. Lett.* **2000**, *15*, 1737–1739. (m) Dudley, D. A.; Bunker, A. M.; Chi, L.; Cody, W. L.; Holland, D. R.; Ignasiak, D. P.; Janiczek-Dolphin, N.; McClanahan, T. B.; Mertz, T. E.; Narasimhan, L. S.; Rapundalo, S. T.; Trautschold, J. A.; van Huis, C. A.; Edmunds, J. J. Rational Design, Synthesis, and Biological Activity of Benzoxazinones as Novel Factor Xa Inhibitors. *J. Med. Chem.* **2000**, *43*, 4063–4070. (n) Masters, J. J.; Franciskovich, J. B.; Tinsley, J. M.; Campbell, C.; Campbell, J. B.; Craft, T. J.; Froelich, L. L.; Gifford-Moore, D. S.; Hay, L. A.; Herron, D. K.; Klimkowski, V. J.; Kurz, K. D.; Metz, J. T.; Ratz, A. M.; Shuman, R. T.; Smith, G. F.; Smith, T.; Towner, R. D.; Wiley, M. R.; Wilson, A.; Yee, Y. K. Non-Amidine-Containing 1,2-Dibenzamidobenzene Inhibitors of Human Factor Xa with Potent Anticoagulant and Antithrombotic Activity. *J. Med. Chem.* **2000**, *43*, 2087–2092. (o) Herron, D. K.; Goodson, T.; Wiley, M. R.; Weir, L. C.; Kyle, J. A.; Yee, Y. K.; Tebbe, A. L.; Tinsley, J. M.; Mendel, D.; Masters, J. J.; Franciskovich, J. B.; Sawyer, J. S.; Beight, D. W.; Ratz, A. M.; Milot, G.; Hall, S. E.; Klimkowski, V. J.; Wikel, J. H.; Eastwood, B. J.; Towner, R. D.; Gifford-Moore, D. S.; Craft, T. J.; Smith, G. F. 1,2-Dibenzamidobenzene Inhibitors of Human Factor Xa. *J. Med. Chem.* **2000**, *43*, 859–872. (p) Wiley, M. R.; Weir, L. C.; Briggs, S.; Bryan, N. A.; Buben, J.; Campbell, C.; Chirgadze, N. Y.; Conrad, R. C.; Craft, T. J.; Ficorilli, J. V.; Franciskovich, J. B.; Froelich, L. L.; Gifford-Moore, D. S.; Goodson, T.; Herron, D. K.; Klimkowski, V. J.; Kurz, K. D.; Kyle, J. A.; Masters, J. J.; Ratz, A. M.; Milot, G.; Shuman, R. T.; Smith, T.; Smith, G. F.; Tebbe, A. L.; Tinsley, J. M.; Towner, R. D.; Wilson, A.; Yee, Y. K. Structure-Based Design of Potent, Amidine-Derived Inhibitors of Factor Xa: Evaluation of Selectivity, Anticoagulant Activity, and Antithrombotic Activity. *J. Med. Chem.* **2000**, *43*, 883–899. (q) Yee, Y. K.; Tebbe, A. L.; Linebarger, J. H.; Beight, D. W.; Craft, T. J.; Gifford-Moore, D.; Goodson, T.; Herron, D. K.; Klimkowski, V. J.; Kyle, J. A.; Sawyer, J. S.; Smith, G. F.; Tinsley, J. M.; Towner, R. D.; Weir, L.; Wiley, M. R. N2-Aroylanthranilamide Inhibitors of Human Factor Xa. *J. Med. Chem.* **2000**, *43*, 873–882.

(17) (a) Ostrem, J. A.; Al-Obeidi, F. A.; Safar, P.; Safarova, A.; Stringer, S. K.; Patek, M.; Cross, M. T.; Spoonamore, J.; LoCascio, J. C.; Kasireddy, P.; Thorpe, D.; Sepetov, N.; Lebl, M.; Wildgoose, P.; Strop, P. Discovery of a novel, potent, and specific family of factor Xa inhibitors via combinatorial chemistry. *Biochemistry* **1998**, *37*, 1053–1059. (b) Mochalkin, I.; Tulinsky, A. Structures of thrombin retro-inhibited with SEL2711 and SEL2770 as they relate to factor Xa binding. *Acta Crystallogr.* **1999**, *D55*, 785–793.

(18) Padmanabhan, K.; Padmanabhan, K. P.; Tulinsky, A.; Park, C. H.; Bode, W.; Huber, R.; Blankenship, D. T.; Cardin, A. D.; Kisiel, W. Structure of human Des(1–45) factor Xa at 2.2 Å resolution. *J. Mol. Biol.* **1993**, *232*, 947–966.

(19) (a) Maignan, S.; Guilloteau, J.-P.; Pouzieux, S.; Choi-Sledeski, Y. M.; Becker, M. R.; Klein, S. I.; Ewing, W. R.; Pauls, H. W.; Spada, A. P.; Mikol, V. Crystal structures of human factor Xa complexed with potent inhibitors. *J. Med. Chem.* **2000**, *43*, 3226–3232. (b) Kamata, K.; Kawamoto, H.; Honma, T.; Iwama, T.; Kim, S.-H. Structural basis for chemical inhibition of human blood coagulation factor Xa. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6630–6635. (c) Brandstetter, H.; Kühne, A.; Bode, W.; Huber, R.; von der Saal, W.; Withensohn, K.; Engh, R. A. X-ray structure of active site-inhibited clotting factor Xa. Implications for drug design and substrate recognition. *J. Biol. Chem.* **1996**, *271*, 29988–29992. (d) Nar, H.; Bauer, M.; Schmid, A.; Stassen, J.-

M.; Wienen, W.; Priepke, H. W. M.; Kauffmann, I. K.; Ries, U. J.; Hauel, N. H. Structural Basis for Inhibition Promiscuity of Dual Specific Thrombin and Factor Xa Blood Coagulation Inhibitors. *Structure* **2001**, *9*, 29–37. (e) Adler, M.; Davey, D. D.; Phillips, G. B.; Kim, S.-H.; Jancarik, J.; Rumennik, G.; Light, D. R.; Whitlow, M. Preparation, Characterization, and the Crystal Structure of the Inhibitor ZK-807834 (CI-1031) Complexed with Factor Xa. *Biochemistry* **2000**, *39*, 12534–12542. (f) Wei, A.; Alexander, R. S.; Duke, J.; Ross, H.; Rosenfeld, S. A.; Chang, C.-H. Unexpected Binding Mode of Tick Anticoagulant Peptide Complexed to Bovine Factor Xa. *J. Mol. Biol.* **1998**, *283*, 147–154.

(20) Defossa, E.; Heinelt, U.; Klingler, O.; Zoller, G.; Al-Obeidi, F.; Walser, A.; Wildgoose, P.; Matter, H. PCT Int. Appl. WO 9933800 A1 199990708, 1999 (199 pages).

(21) Dougherty, D. A. Cation–π interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. *Science* **1996**, *271*, 163–168.

(22) Heinelt, U.; Herok, S.; Matter, H.; Wildgoose, P. Solid-phase optimisation of achiral amidinobenzyl indols as potent and selective factor Xa inhibitors. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 227–230.

(23) Kroemer, R. T.; Ettmayer, P.; Hecht, P. 3D-Quantitative Structure–Activity Relationships of Human Immunodeficiency Virus Type-1 Protein as Inhibitors: Comparative Molecular Field Analysis of 2-Heterosubstituted Statine Derivatives. Implication for the Design of Novel Inhibitors. *J. Med. Chem.* **1995**, *38*, 4917–4928.

(24) PDB files Protein Database (National Brookhaven Laboratories): 1HCG, 1XKA, 1FAX (http://www.rcsb.org/pdb). (a) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535–542. (b) Berman, J. W.; Zukang, F.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(25) Cramer, R. D.; Patterson, D. E.; Bunce, J. E. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.

(26) Clark, M.; Cramer, R. D.; Jones, D. M.; Patterson, D. E.; Simeroth, P. E. Comparative Molecular Field Analysis (CoMFA). 2. Towards its Use with 3D-Structural Databases. *Tetrahedron Comput. Methodol.* **1990**, *3*, 47–59.

(27) *3D-QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993.

(28) Klebe, G.; Abraham, U.; Mietzner, T. Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules To Correlate and Predict Their Biological Activity. *J. Med. Chem.* **1994**, *37*, 4130–4146.

(29) (a) Wold, S.; Albano, C.; Dunn, W. J.; Edlund, U.; Esbenson, K.; Geladi, P.; Hellberg, S.; Lindberg, W.; Sjöström, M. In *Chemometrics: Mathematics and Statistics in Chemistry*; Kowalski, B., Ed., Reidel: Dortrecht, The Netherlands, 1984; pp 17–95. (b) Dunn, W. J.; Wold, S.; Edlund, U.; Hellberg, S.; Gasteiger, J. Multivariate Structure–Activity Relationship between Data from a Battery of Biological Tests and an Ensemble of Structure Descriptors: The PLS Method. *Quant. Struct.–Act. Relat.* **1984**, *3*, 31–137. (c) Geladi, P. Notes on the History and Nature of Partial Least Squares (PLS) Modelling. *J. Chemom.* **1988**, *2*, 231–246.

(30) (a) Wold, S. Cross-Validatory Estimation of the Number of Components in Factor and Principal Component Models. *Technometrics* **1978**, *4*, 397–405. (b) Diaconis, P.; Efron, B. Computer-Intensive Methods for Statistics. *Sci. Am.* **1984**, *116*, 96–117. (c) Cramer, R. D.; Bunce, J. D.; Patterson, D. E. Crossvalidation, Bootstrapping and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies. *Quant. Struct.–Act. Relat.* **1988**, *7*, 18–25.

(31) Moss, R. A.; Ma, W.; Merrer, D. C.; Xue, S. Conversion of "Obstinate" Nitriles to Amides by Garigipati's Reaction. *Tetrahedron Lett.* **1995**, *36*, 8761–8764.

(32) Ramasamy, K.; Robins, R. K.; Revankar, G. R. A Convenient Synthesis of 5-Substituted-7-β-D-arabinofuranosylpyrrolo[2,3-d]-pyrimidines Structurally Related to the Antibiotics Toyocamycin and Sangivamycin. *J. Heterocycl. Chem.* **1988**, *25*, 1043–1046.

(33) Wagner, G.; Voigt, B.; Vieweg, H. Synthesis of Nα-(arylsulfonylglycyl)amidinophenylalaninamides as highly active inhibitors of thrombin. *Pharmazie* **1984**, *39*, 226–230.

(34) Kabsch, W. Evaluation of single-crystal x-ray diffraction data from a position sensitive detector. *J. Appl. Crystallogr.* **1988**, *21*, 916–924.

(35) Brünger, A. T. *X-PLOR: A System for X-ray Crystallography and NMR*, version 3.1; Yale University Press: New Haven, CT, 1992.

(36) *SYBYL Molecular Modelling Package*, version 6.6; Tripos: St. Louis, MO, 2000.

(37) (a) Halgren, T. Maximally diagonal force constants in dependent angle-bending coordinates. 2. Implications for the design of empirical force fields. *J. Am. Chem. Soc.* **1990**, *112*, 4710−4723. (b) Halgren, T. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **1999**, *20*, 720−729.

(38) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.

(39) McMartin, C.; Bohacek, R. S. QXP: powerful, rapid computer algorithms for structure-based drug design. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 333−344.

(40) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.

(41) Kearsley, S. K.; Smith, G. M. An alternative method for the alignment of molecular structures: Maximizing electrostatic and steric overlap. *Tetrahedron Comput. Methodol.* **1990**, *3*, 615−633.

(42) (a) Thibaut, U.; Folkers, G.; Klebe, G.; Kubinyi, H.; Merz, A.; Rognan, D. Recommendations for CoMFA studies and 3D QSAR publications. In *3D QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993; pp 711−717. (b) Folkers, G.; Merz, A.; Rognan, D. CoMFA: Scope and limitations. In *3D QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993; pp 583−616. (c) Cramer, R. D.; DePriest, S. A.; Patterson, D. E.; Hecht, P. The developing practice of comparative molecular field analysis. In *3D QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993; pp 443−485.

(43) Sheridan, R. P.; Nachbar, R. B.; Bush, B. L. Extending the trend vector: The trend matrix and sample-based partial least squares. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 323−340.

(44) Clark, M.; Cramer, R. D. The probability of chance correlation using partial least squares (PLS). *Quant. Struct.−Act. Relat.* **1993**, *12*, 137−145.

(45) Leonard, J. M.; Clark, R. D. PLS Validation by progressive scrambling. Presented at the 12th European Symposium on Quantitative Structure−Activity Relationships (QSAR'98), 1998; poster.

(46) *UNITY Chemical Information Software*, version 4.1; Tripos Inc., 1699 S. Hanley Road, St. Louis, MO 63144.

(47) (a) Willett, P. *Similarity and Clustering in Chemical Information Systems*; Research Studies Press: Letchworth, U.K., 1987. (b) Willett, P.; Winterman, V. A. Comparison of Some Measures for the Determination of Intermolecular Structural Similarity. *Quant. Struct.−Act. Relat.* **1986**, *5*, 18−25.

(48) Lajiness, M.; Johnson, M. A.; Maggiora, G. M. Implementing Drug Screening Programs Using Molecular Similarity Methods. In *QSAR: Quantitative Structure−Activity Relationships in Drug Design*; Fauchere, J. L., Ed.; Alan R. Liss Inc.: New York, 1989; pp 173−176.

(49) Taylor, R. Simulation Analysis of Experimental Design Strategies for Screening Random Compounds as Potential New Drugs and Agrochemicals. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 59−67.

(50) Biological activities are expressed as $\log[(1/K_i)100000]$.

(51) Pötter, T.; Matter, H. Random or Rational Design? Evaluation of Diverse Compound Subsets from Chemical Structure Databases. *J. Med. Chem.* **1998**, *41*, 478−488.

(52) Matter, H.; Kotsonis, P.; Klingler, O.; Strobel, H.; Fröhlich, L.; Frey, A.; Pfleiderer, W.; Schmidt, H. H. H. W. Structural Requirements for Inhibition of the Neuronal Nitric Oxide Synthase (NOS-I): 3D-QSAR Analysis of 4-Oxo- and 4-Amino-pteridine Based Inhibitors. *J. Med. Chem.*, in press.

(53) (a) Klebe, G.; Abraham, U. Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 1−10. (b) Böhm, M.; Stürzebecher, J.; Klebe, G. Three-Dimensional Quantitative Structure−Activity Relationship Analyses Using Comparative Molecular Field Analysis and Comparative Molecular Similarity Indices Analysis To Elucidate Selectivity Differences of Inhibitors Binding to Trypsin, Thrombin, and Factor Xa. *J. Med. Chem.* **1999**, *42*, 458−477. (c) Klebe, G. Comparative molecular similarity indices analysis. CoMSIA. *Perspect. Drug Discovery Des.* **1998**, *12−14*, 87−104.

(54) ACD (Available Chemicals Directory) from Molecular Design Ltd., 14600 Catalina Street, San Leandro, CA 94577.

(55) (a) Gallivan, J. P.; Dougherty, D. A. A Computational Study of Cation−π Interactions vs Salt Bridges in Aqueous Media: Implications for Protein Engineering. *J. Am. Chem. Soc.* **2000**, *122*, 870−874. (b) Gallivan, J. P.; Dougherty, D. A. Cation−π interactions in structural biology. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9459−9464. (c) Ma, J. C.; Dougherty, D. A. The Cation−π Interaction. *Chem. Rev.* **1997**, *97*, 1303−1324.

(56) Monnaie, D.; Arosio, D.; Griffon, N.; Rose, T.; Rezaie, A. R.; Di Cera, E. Identification of a Binding Site for Quertary Amines in Factor Xa. *Biochemistry* **2000**, *39*, 5349−5354.

(57) Flocco, M. M.; Mowbray, S. L. Planar stacking interactions of arginine and aromatic side-chains in proteins. *J. Mol. Biol.* **1994**, *235*, 709−717.

(58) Sulpizi, M.; Carloni, P. Cation−π versus OH−π Interactions in Proteins: A Density Functional Study. *J. Chem. Phys. B* **2000**, *104*, 10087−10091.

(59) Kastenholz, M. A.; Pastor, M.; Cruciani, G.; Haaksma, E. E. J.; Fox, T. GRID/CPCA: A New Computational Tool To Design Selective Ligands. *J. Med. Chem.* **2000**, *43*, 3033−3044.

(60) Matter, H.; Defossa, E.; Heinelt, U.; Naumann, T.; Schreuder, H.; Wildgoose, P. Combining Structure-Based Design and 3D-QSAR towards the Discovery of Non-Chiral, Potent and Selective Factor Xa Inhibitors. In *Rational Approaches to Drug Design, Proceedings of the 13th European Symposium on Quantitative Structure−Activity Relationships*; Höltje, H.-D., Sippl, W., Eds.; Prous Science: Barcelona, Spain, 2001; pp 177−185.

(61) Manuscript in preparation.

(62) Matter, H.; Schwab, W.; Barbier, D.; Billen, G.; Haase, B.; Neises, B.; Schudok, M.; Thorwart, W.; Schreuder, H.; Brachvogel, V.; Loenze, P.; Weithmann, K. U. Quantitative Structure−Activity Relationship of Human Neutrophil Collagenase (MMP-8) Inhibitors Using Comparative Molecular Field Analysis and X-ray Structure Analysis. *J. Med. Chem.* **1999**, *42*, 1908−1920.

JM0111346