Bioorganic & Medicinal Chemistry 20 (2012) 5453–5460

Contents lists available at SciVerse ScienceDirect

Bioorganic & Medicinal Chemistry

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

Who cares for the protons?

Paul Czodrowski

Merck KGaA, Computational Chemistry, Frankfurter Strasse 250, D-64293 Darmstadt, Germany

ARTICLE INFO

Article history:
Available online 10 March 2012

Keywords:
pK_a calculation
Drug design
Protonation
BIPS

ABSTRACT

It is tempting to use standard protonation states for the analysis of protein–ligand interactions. Two different pK_a calculation methods, PROPKA (protein pK_a) and MCCE (multi conformation continuum electrostatics), were applied to challenge this convenient behavior. As data basis, we selected five recently approved drugs for which structural information of the protein–drug complex is available. We analyzed the pK_a calculations in terms of a measure termed *BIPS* (binary protonation states) recently introduced by us. Both methods agree in detecting the majority of the sites with atypical *BIPS* values. However, when using only one method, some of the atypcial *BIPS* value would have been missed. Therefore, we recommend using both methods to set such an interpretation on a solid basis.

1. Introduction

With the help of protein pK_a calculations, one is able to estimate the protonation states of proteins and protein–ligand complexes. This is relevant for an understanding of molecular recognition. It may also add an often overlooked effect to the interpretation of structure–activity relationships of ligands, respectively drugs or drug candidates, ultimately leading to better rational design of drugs.

Recent benchmarks on state of the art pK_a calculation algorithms revealed that the two programs PROPKA[1] (protein pK_a) and MCCE[2] (multi conformation continuum electrostatics) obtain the best results.[3,4] These two program use different methodologies: PROPKA makes use of empirically derived rules based on experimental data. In contrast, MCCE utilizes a continuum solvent model that employs the Poisson–Boltzmann equation.

In this contribution, we perform a stress-test of these two methods for a rather small but relevant dataset. Additionally, we investigate the effect of conformational sampling in the MCCE experiments. It is known that pK_a calculations perform best for surface residues with only small pK_a shifts, whereas accuracy breaks down for highly buried residues usually found in active sites.[3] Therefore, we are eager to examine their performance on the active site of protein–ligand complexes of pharmaceutical relevance. From the 2011 FDA approvals, we selected five drugs with publically available co-crystal structures. By using such a selection, we want to evaluate the potential for pK_a calculations in an industrialized drug design framework. Furthermore, the target space depicts the structural playground of structure-based drug design of the recent years: Besides two serine proteases (Factor Xa and DPP IV), one phosphodiesterase (PDE4D) is found, as well as one protein tyrosine kinase (RET kinase), and lastly the reverse transcriptase of the human immunodeficiency virus type 1 (HIV-RT). A similar statement regarding structural diversity can be given for the investigated drugs: a 2D depiction of the drugs in their binding pockets can be found in Figure 1.

Primarily, we will evaluate our pK_a calculations in terms of a measure we recently introduced,[5] known as *BIPS* (binary protonation state). This measure translates a pK_a value into a discrete value for a protonation state, that is only fully protonated or fully de-protonated residues are considered. This view reflects the necessity when dealing with protein–ligand complexes for post-processing with molecular modeling tools. Here, only discrete values can be adjusted for the titratable group, so that a pK_a value needs to be transformed to a protonation state.

2. Materials and methods

The protein preparation wizard of the Maestro suite was used for analysis and preparation of the proteins.[6] The following protein structures from the Protein Data Bank were used: 2RGU (dipeptidyl peptidase IV or DPP IV),[7] 2W26 (Factor Xa),[8] 2ZD1 (HIV-1 reverse transcriptase or HIV-RT),[9] 1XOQ (phosphodiesterase 4D or PDE4D),[10] 2IVU (rearranged during transfection gene coding for RET tyrosine kinase).[11] If there were multiple structures of the same protein found in the PDB, the structure with the highest resolution was chosen. In all structures, any crystal buffer molecule was deleted prior to the calculation. The crystal water molecules were kept for the pK_a calculations by MCCE. For DPP IV and HIV-RT, the protein chain A was used for the pK_a calculations. Although DPP IV and HIV-RT are not active in their monomeric form, this decision was driven by the computational burden. The RET tyrosine kinase crystal structure contains two gaps: GLU-713 is missing, as is the loop between the residues LYS-812 and GLU-843. The side chain of GLU-713 was re-built using PRIME[12] of the

E-mail address: paul.czodrowski@merckgroup.com

0968-0896/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.bmc.2012.03.009

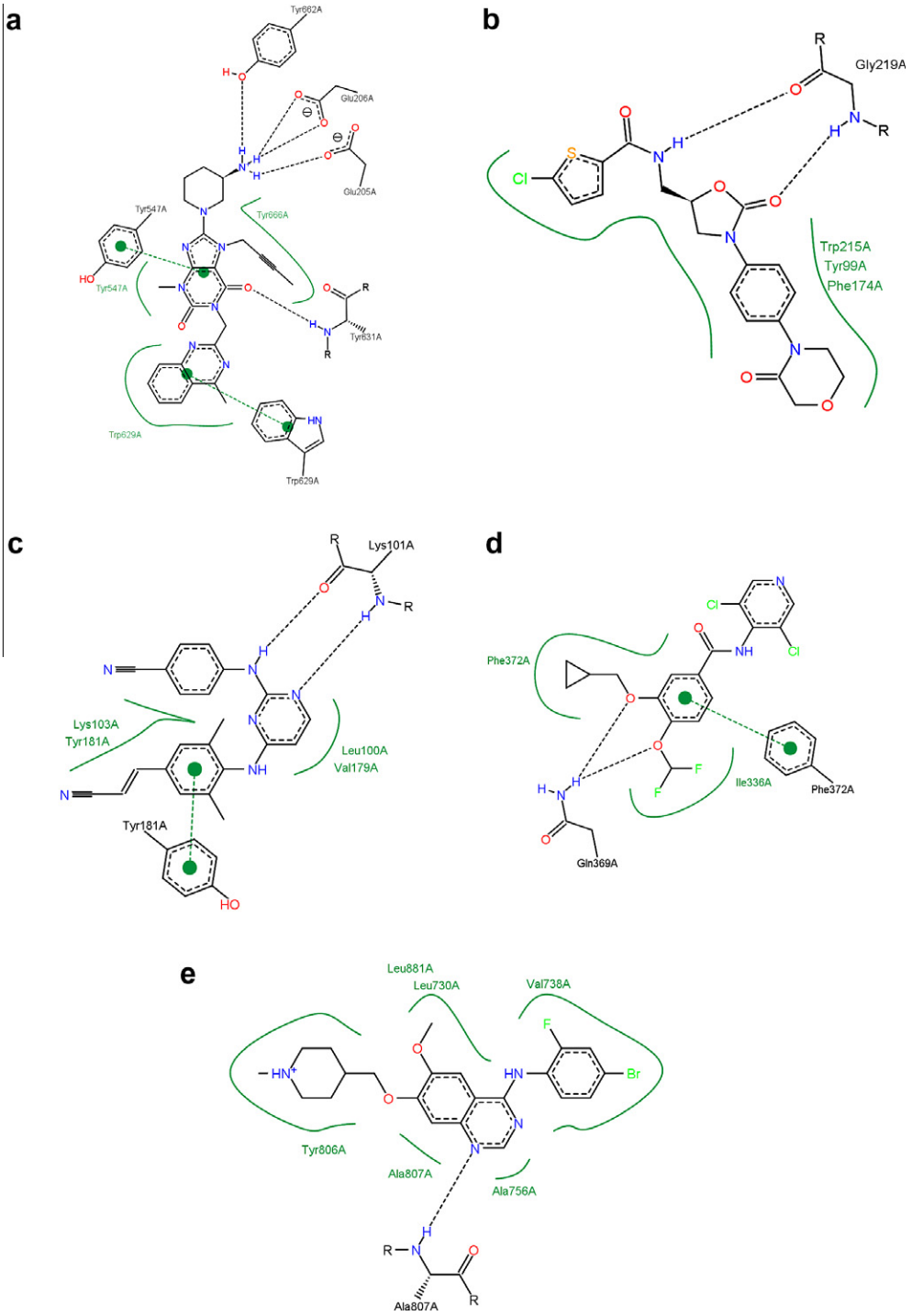

Figure 1. 2D depiction (generated by PoseView[30]) of the investigated drugs: (a) linagliptin bound to DPP IV (b) rivaroxaban bound to factor Xa (c) rilpivirine bound to HIV-RT (d) roflumilast bound to PDE4D (e) vandetanib bound to RET tyrosine kinase. Pictures have been generated by PoseView (http://www.poseview.de).

Schrödinger suite. The missing loop was not rebuilt, and the 'terminal' residues were capped with neutralizing groups.

Version 2.7 of MCCE,[2] and version 3.1 of PROPKA[1] were used. No parameters needed to be set for PROPKA calculations. The rest of this paragraph therefore only refers to MCCE. Two different techniques for conformation sampling were used: DEFAULT (run.prm.default) and QUICK (run.prm.quick). On average, 2.5 conformers per titratable residue are sampled in the QUICK runs, whereas 13.5 conformers are sampled in the DEFAULT MCCE runs.[2] PEOE_PB charges were used for the ligand,[13] while AMBER charges were used for the protein.[14]

The model pK_a value of the two drugs that have titratable groups were calculated with CompuDrug pKalc in the version 4.3.0.[15] The computed ligand pK_a values for linagliptin and vandetanib are 10.15 and 9.16, respectively.

3. Theory

A pK_a value of a titratable group in a protein or a ligand bound to a protein can be described by a thermodynamic cycle.[16] This basic assumption is common to all protein pK_a calculation algorithms, and the following reactions are taken into account:

- The acid–base equilibrium can be described by the reaction in aqueous solution.
- The acid and base is transferred from the aqueous solution to the protein interior.

The computed pK_a values are split into different terms to denote this thermodynamic cycle. Each titratable group has a *model pK_a value* reflecting its pK_a value in aqueous solution. This represents the former point of the above listing. The desolvation of the titratable group into the protein and the electrostatic interaction to the permanent dipoles gives the *intrinsic pK_a value*.[16] This represents the latter point of the above listing. Finally, the influence of the other titratable groups—termed as *site–site interactions*—gives the computed pK_a value.

What is special about continuum-electrostatics methods such as MCCE is the choice of the *protein dielectric constant* (ε). It is a central point of discussion.[17] In our publication, we use the term 'eps' to abbreviate the dielectric constant. To probe its influence, two different values (4 and 8) are used in our MCCE calculations.

Strongly coupled electrostatic interactions between different titratable groups can give rise to *perturbed titration curves*.[18,19] Their curvature is significantly different from a sigmoidal shape. Usually, this behavior can be rationalized through a detailed analysis of the titration partners. Experimentally, such curves are also observed.[20–22] Such an analysis is possible when using MCCE. Although PROPKA does not output any titration curve, the most recent version is capable of detecting strongly coupled groups.[1]

3.1. BIPS

The protonation state of a titratable group is calculated using the Henderson–Hasselbalch equation:

$$\mathrm{pH} = \mathrm{p}K_a - \log\frac{[\mathrm{deprotonated form}]}{[\mathrm{protonated\ form}]} \quad (1)$$

If one assumes the condition [deprotonated form] + [protonated form] = 1, the protonation state can be calculated by a rearrangement of the Henderson–Hasselbalch equation:

$$\text{protonation state} = \frac{1}{10^{(\mathrm{pH}-\mathrm{p}K_a)}+1} \quad (2)$$

In this publication, the pH is set to a value of 7.4 in order to represent physiological conditions. The *BIPS* value is now set according to the computed protonation state:

$$\begin{aligned} BIPS = 0 \quad & \begin{cases} \text{Any acidic group : protonation state } < 0.1 \\ \text{Any basic group : protonation state } > 0.9 \end{cases} \\ \underset{[aBIPS]}{BIPS} = 1 \quad & \begin{cases} \text{Any acidic group : protonation state } \geq 0.1 \\ \text{Any basic group : protonation state } \leq 0.9 \end{cases} \end{aligned} \quad (3)$$

A standard protonation state (protonated base and de-protonated acid) refers to a BIPS value of 0. In analogy, a non-standard *BIPS* (*BIPS* = 1) value refers to a de-protonated base and a protonated acid. Although histidines belong to the basic group, they are considered in a different manner. They have an atypical *BIPS* value, that is *BIPS* = 1, if they are protonated by more than 90%. Atypical *BIPS* values (protonated acids and deprotonated bases), that is *BIPS* = 1, are annotated as *aBIPS*.

4. Results

The workflow for the analysis of the pK_a calculations is as follows:

- The residues with *aBIPS* values are counted.
- The overlap of the PROPKA and MCCE calculations is analyzed.
- Residues with *aBIPS* values outside the active site are briefly discussed.
- A detailed analysis of residues with *aBIPS* values is done for residues in a radius of 6 Å around the ligand.

The emphasis of our analysis towards a 6 Å sphere around the ligand is motivated by the pragmatism toward putative ligand design. All results (residues with *aBIPS* values and the corresponding pK_a values) can be found throughout the Tables 1–5.

4.1. DPP IV (linagliptin—Tradjenta™) Table 1

The PROPKA calculation gives 9 and 10 residues with *aBIPS* values. These numbers are much lower in the MCCE calculations: here, between 2 and 6 residues are classified with *aBIPS* values. An overlap between the two programs is found for the N-terminus and ASP-230. Both residues are 28 Å and 15 Å remote to the ligand. In the MCCE calculations, the site–site interactions are responsible for the *aBIPS* value of the N-terminus. In the PROPKA calculations, this residue is almost unshifted. Here, the *aBIPS* value is caused by the rather low model pK_a value (8.0) of the N-terminus. In the case of ASP-230, the desolvation and the site–site interaction to ASP-200 are responsible for the *aBIPS* value in the MCCE calculation. In the PROPKA calculation, all pK_a modulating terms (desolvation, sidechain hydrogen bond, backbone hydrogen bond, coulombic interaction) have an upwards shift for ASP-230.

The PROPKA calculation gives a *aBIPS* for ASP-663 (apo and complexed) which is located in the active site. Desolvation and the interaction to GLU-205 and GLU-206 are the major driving forces for this upwards shift. When complexed with linagliptin, the ligand has a lowering effect on the pK_a value of ASP-663. However, the *aBIPS* value is still found for ASP-663. The MCCE calculations reveal that ASP-663 is strongly coupled to GLU-205 and GLU-206. For one member of the catalytic triad, namely HIS-740, a consistent protonation trend is observed in the MCCE calculations. In the apo form, this residue shows a pK_a value with an upwards shift compared to its model pK_a value. In most, but not all cases, this results in a perturbed *BIPS* value. This is due to the fact that its value is at the borderline of a *BIPS* transition. In the complexed form, this residue shows a lowered pK_a value—compared to the value of the apo form—resulting in a normal *BIPS* value. HIS-740 is located between the pyrimidinophenyl group of the ligand linagliptin and the other two members of the catalytic triad, ASP-708 and SER-630.

4.2. Faxtor Xa (Rivaroxaban—Xarelto™) Table 2

The PROPKA calculation only gives one *aBIPS* value, whereas the MCCE calculations reveal between 1 and 5 *aBIPS* values. Both methods are aligned in deciphering the *aBIPS* value of the N-terminus. The other *aBIPS* values in the region outside the active site are one residue pair close in space: TYR-60 and LYS-90. The calculation reveals that these residues have a charge-charge interaction. Although only the QUICK/ε = 4 calculation gives a 'true' *aBIPS* value for LYS-90, the other MCCE calculations also display a strongly shifted pK_a value for this residue. In the active site, ASP-189 emerges with an *aBIPS* value, but in complex with rivaroxaban. In the apo form, ASP-189 shows a normal *BIPS* value. The distance between OD1 of ASP-189 and the thiophene chlorine atom of rivaroxaban amounts to 4.7 Å. For three out of four MCCE calculations, an *aBIPS* value is reported for ASP-189 in the complexed form. The DEFAULT-MCCE/eps = 8 calculation gives a pK_a value for ASP-189 (6.23) which is not high enough to be considered as *aBIPS*, since it is protonated by only 6%. To be considered as *aBIPS*, 10% protonation would be necessary. The PROPKA calculation also reveals a normal *BIPS* value for ASP-189: it remains almost unchanged going

Table 1
Residues giving aBIPS values for apo DPP IV and the linagliptin/DPP IV complex. The corresponding pK_a value follows in the second line

DEFAULT-MCCE/eps = 4		DEFAULT-MCCE/eps = 8		QUICK-MCCE/eps = 4		QUICK-MCCE/eps = 8		PROPKA	
apo	cpl	apo	cpl	apo	cpl	apo	cpl	apo	cpl
NTR39	NTR39	NTR39	NTR39	NTR39	NTR39	NTR39	NTR39	NTR39	NTR39
pK_a = 5.44	pK_a = 5.34	pK_a = 6.51	pK_a = 6.48	pK_a = 3.39	pK_a = 3.40	pK_a = 5.07	pK_a = 5.07	pK_a = 7.35	pK_a = 7.35
				GLU117	GLU117	GLU117	GLU117		
				pK_a = 8.04	pK_a = 8.06	pK_a = 6.80	pK_a = 6.74		
						ASP200			
						pK_a = 6.51			
								GLU204	GLU204
								pK_a = 7.18	pK_a = 7.48
						ASP206.x			
						pK_a = 8.36			
		ASP230	ASP230			ASP230	ASP230	ASP230	ASP230
		pK_a = 7.01	pK_a = 6.94			pK_a = 11.38	pK_a = 11.60	pK_a = 11.53	pK_a = 11.53
								LYS258	LYS258
								pK_a = 7.35	pK_a = 7.27
								ASP302	ASP302
								pK_a = 7.30	pK_a = 7.30
ASP326	ASP326			ASP326	ASP326				
pK_a = 7.08	pK_a = 7.00			pK_a = 7.02	pK_a = 7.22				
				GLU361	GLU361				
				pK_a = 6.70	pK_a = 6.50				
								LYS512	LYS512
								pK_a = 8.08	pK_a = 8.05
								ASP579	ASP579
								pK_a = 7.19	pK_a = 7.23
								ASP663.x	ASP663.x
								pK_a = 11.77	pK_a = 10.92
									ASP708
									pK_a = 7.55
								ASP709	ASP709
								pK_a = 6.72	pK_a = 7.38
		HIS740.x				HIS740.x			
		pK_a = 9.14				pK_a = 8.73			
	LIG: pK_a >14		LIG: pK_a >14		LIG: pK_a >14		LIG:		LIG: pK_a 11.9
	[100% prot]		[100% prot]		[100% prot]		pK_a >14[100%		[100% prot]
							prot]		

The suffix (.x) denotes a distance between this residue and the ligand smaller than 6 Å.

Table 2
Residues giving aBIPS values for apo Factor Xa and the rivaroxaban/Factor Xa complex. The corresponding pK_a value follows in the second line

DEFAULT-MCCE/eps = 4		DEFAULT-MCCE/eps = 8		QUICK MCCE/eps = 4		QUICKMCCE/eps = 8		PROPKA	
apo	cpl	apo	cpl	apo	cpl	apo	cpl	apo	cpl
NTE-B	NTE-B	NTE-B	NTE-B	NTE-B	NTE-B	NTE-B	NTE-B	NTE-B	NTE-B
pK_a = 5.83	pK_a = 5.82	pK_a = 6.81	pK_a = 6.81	pK_a = 5.33	pK_a = 5.34	pK_a = 6.34	pK_a = 6.35	pK_a = 7.36	pK_a = 7.18
TYR-60	TYR-60	TYR-60	TYR-60	TYR-60	TYR-60				
pK_a = 6.03	pK_a = 6.06	pK_a = 7.11	pK_a = 7.13	pK_a = 7.90	pK_a = 7.77				
GLU-80	GLU-80		GLU-80	GLU-80	GLU-80		GLU-80		
pK_a = 7.99	pK_a = 12.43		pK_a = 8.61	pK_a = 9.08	pK_a = 12.59		pK_a = 9.06		
				LYS-90	LYS-90				
				pK_a = 7.53	pK_a = 7.51				
	ASP-189.x				ASP-189.x		ASP-189.x		
	pK_a = 8.87				pK_a = 10.35		pK_a = 7.06		

The suffix (.x) denotes a distance between this residue and the ligand smaller than 6 Å.

from the apo to the complexed form. It shows a pK_a value of 6.34 and 6.33, respectively.

4.3. HIV-RT (rilpivirine—Edurant™) Table 3

The PROPKA calculations identify two residues with *aBIPS* values (apo and holo form), whereas the MCCE calculations classify between 1 and 10 residues bearing atypical *BIPS* values. There is a partial intersection between the two calculation methods. The QUICK-MCCE calculations agree in deciphering *aBIPS* value for ASP-443 and GLU-546 as found by PROPKA. Both residues are more than 60 Å distant from the ligand. However, both residues are close in space to each other: the distance between OD2 (ASP-443) and OE2 (GLU-546) amounts to 4.4 Å. In this region, the RNase H subdomain is bound to HIV-RT. None of the active site residues shows an *aBIPS* value. Mutation studies reveal the relevance of one member of the active site (LYS-103): in 57% of the cases when a non-nucleoside reverse transcriptase inhibitors is bound to HIV-RT, LYS-103 is mutated to asparagine.[23,24] This mutation maintains the positive formal charge of the residue which is found by the pK_a calculations as well: all calculations agree in assigning a normal *BIPS* value to LYS-103.

Table 3
Residues giving aBIPS values for apo HIV-RT and the rilpivirine/HIV-RT complex. The corresponding pK_a value follows in the second line

DEFAULT-MCCE/eps = 4		DEFAULT-MCCE/eps = 8		QUICK-MCCE/eps = 4		QUICK-MCCE/eps = 8		PROPKA	
apo	cpl	apo	cpl	apo	cpl	apo	cpl	apo	cpl
		NTR	NTR	NTR	NTR	NTR	NTR		
		pK_a = 8.25	pK_a = 8.24	pK_a = 7.98	pK_a = 7.98	pK_a = 7.79	pK_a = 7.79		
				LYS65	LYS65				
				pK_a = 8.31	pK_a = 8.31				
LYS73	LYS73			LYS73	LYS73	LYS73	LYS73		
pK_a = 8.01	pK_a = 8.05			pK_a = 7.26	pK_a = 7.21	pK_a = 8.34	pK_a = 8.21		
					GLU89				
					pK_a = 6.47				
				LYS281	LYS281	LYS281	LYS281		
				pK_a = 7.43	pK_a = 7.41	pK_a = 8.17	pK_a = 8.15		
	ASP443			ASP443	*ASP443*	ASP443	ASP443	ASP443	ASP443
	pK_a = 12.81			pK_a >14	pK_a >14	pK_a = 10.12	pK_a = 10.14	pK_a = 7.06	pK_a = 7.06
				LYS454	LYS454				
				pK_a = 7.49	pK_a = 7.50				
				ASP498	ASP498				
				pK_a = 8.30	pK_a = 8.47				
				LYS512	LYS512	LYS512	LYS512		
				pK_a = 6.09	pK_a = 6.06	pK_a = 7.74	pK_a = 7.72		
				GLU546	GLU546	GLU546	GLU546	GLU546	GLU546
				pK_a >14	pK_a >14	pK_a >14	pK_a >14	pK_a = 8.33	pK_a = 8.33
LYS550	LYS550					LYS550	LYS550		
pK_a = 6.07	pK_a = 6.32					pK_a = 6.76	pK_a = 6.81		

The suffix (.x) denotes a distance between this residue and the ligand smaller than 6 Å.

Table 4
Residues giving aBIPS values for apo PDE4D and the roflumilast/PDE4D complex. The corresponding pK_a value follows in the second line

DEFAULT-MCCE/eps = 4		DEFAULT-MCCE/eps = 8		QUICK-MCCE/eps = 4		QUICK-MCCE/eps = 8		PROPKA	
apo	cpl	apo	cpl	apo	cpl	apo	cpl	apo	cpl
						HIS105			
						pK_a = 8.36			
HIS154	HIS154	HIS154	HIS154	HIS154	HIS154	HIS154	HIS154		
pK_a = 8.63	pK_a = 8.57	pK_a = 8.44	pK_a = 8.52	pK_a = 8.54	pK_a = 8.54sig	pK_a = 8.46	pK_a = 8.43		
HIS160.x		HIS160.x							
pK_a = 7.29		pK_a = 8.39							
HIS233	HIS233	HIS233	HIS233	HIS233	HIS233	HIS233	HIS233		
pK_a >14	pK_a >14	pK_a = 11.38	pK_a = 11.52	pK_a >14	pK_a >14	pK_a = 11.59	pK_a = 12.18		
	TYR303								
	pK_a = 8.15								
	HIS315.x	HIS315.x	HIS315.x		HIS315.x		HIS315.x		
	pK_a = 9.70	pK_a = 8.56	pK_a = 9.36		pK_a = 10.36		pK_a = 9.71		
								GLU339	GLU339
								pK_a = 6.61	pK_a = 6.88
		HIS378	HIS378	HIS378	HIS378	HIS378	HIS378		
		pK_a = 8.52	pK_a = 8.48	pK_a = 8.60	pK_a = 8.56	pK_a = 8.66	pK_a = 8.67		

The suffix (.x) denotes a distance between this residue and the ligand smaller than 6 Å.

Table 5
Residues giving aBIPS values for apo RET tyrosine kinase and the vandetanib/RET tyrosine kinase complex. The corresponding pK_a value follows in the second line

FULL-MCCE/eps = 4		FULL-MCCE/eps = 8		QUICK-MCCE/eps = 4		QUICK-MCCE/eps = 8		PROPKA	
apo	cpl	apo	cpl	apo	cpl	apo	cpl	apo	cpl
ASP874	ASP874			ASP874	ASP874				
pK_a = 7.49	pK_a = 7.4			pK_a = 6.85	pK_a = 6.8				
		PTR905	PTR905			PTR905	PTR905		
		pK_a = 11.18	pK_a = 11.42			pK_a = 10.87	pK_a = 10.85		
LYS907	LYS907	LYS907		LYS907	LYS907	LYS907	LYS907		
pK_a = 8.31	pK_a = 8.03	pK_a = 8.13		pK_a = 7.46	pK_a = 7.42	pK_a = 7.21	pK_a = 7.22		
								GLU943	GLU943
								pK_a = 6.63	pK_a = 6.62
			ZD6				ZD6		
			LIG:				LIG:		
			pK_a = 6.62				pK_a = 6.1		

The suffix (.x) denotes a distance between this residue and the ligand smaller than 6 Å.

4.4. PDE4D (roflumilast—Daliresp™) Table 4

The PROPKA calculations identify one residue with *aBIPS* values (apo and holo form), whereas the MCCE calculations classify between 3 and 5 residues bearing atypical *BIPS* values. No overlap between the PROPKA and MCCE results is observed. GLU-339 is the single residue yielding an *aBIPS* value in the PROPKA calculations. In contrast, the pK_a values of GLU-339 are found to be below 0 are just above 0 in the MCCE calculations. This is caused by site–site interactions not being compensated through desolvation. It is notable that one residue type emerges from the MCCE calculations: in all MCCE calculations, 31 residues (often apo and complexed as well) show *aBIPS* values, and 30 of those belong to histidines. In contrast, PROPKA gives neutral histidines in all cases. A consistent perturbed *BIPS* value (for the apo as well as the complexed form) is given for HIS-154 and HIS-233, both are not members of the active site. HIS-154 is flanked by two aspartates (ASP-156, ASP-203) and is 14 Å distant to the ligand. None of the ASP pK_a values is largely shifted upon ligand binding. HIS-233 is 8 Å distant to the ligand, 7 Å remote to the Zn ion and 4 Å distant to ASP-201. A clear effect of ligand binding is the pK_a signature of HIS-315. Except for the PROPKA and both DEFAULT-MCCE/eps = 8 calculations, HIS-315 shows a perturbed *BIPS* value, indeed only for complexed form. The ligand pyridine ring and the HIS-315 imidazole ring are 5.9 Å distant to each other.

4.5. RET tyrosine kinase (vandetanib—Zactima™) Table 5

In the PROPKA calculations, one *aBIPS* value is found for the apo and the holo form, whereas 1–3 such values are found in the MCCE calculations. There is no intersection between the PROPKA and the MCCE calculations. The structure of the studied RET tyrosine kinase is found in the phosphorylated form. The phosphorylated tyrosine (PTR-905) is 18 Å distant to the drug vandetanib. With a high dielectric constant (eps = 8; DEFAULT and QUICK settings), the phosphorylated tyrosine PTR-905 shows a perturbed *BIPS* value in the MCCE calculations. In the vicinity of PTR-905 another residue with an *aBIPS* value is found, namely LYS-907. PROPKA estimates an *aBIPS* value for GLU-943. It is located in a different region of RET tyrosine kinase, namely the α–F helix, and it is 16 Å remote to the ligand. Finally, the titratable group of vandetanib (methyl-piperidine) shows a perturbed *BIPS* value in the MCCE calculation, but only in those using the higher dielectric constant of 8. PROPKA estimates a normal BIPS value for this titratable group.

5. Discussion

5.1. DPP IV

Both methods agree in detecting coupled systems in the active site, namely ASP-230 and ASP-663. In the case of ASP-230, both methods consistently find its *aBIPS* value. This does not hold true for ASP-663: here, only PROPKA outputs an *aBIPS* value. MCCE seems to de-emphasize the de-protonation effect by considering multiple conformers. Nonetheless, both methods agree in recognizing these 'pK_a hot spots'. PROPKA fails in uncovering the *aBIPS* value of HIS-740 in the uncomplexed state. This is due to the fact the pK_a value in the apo state results in only 80% protonation of this residue, 90% would be necessary to assign an *aBIPS* value. Finally, it must be accentuated that the active site of DPP IV is a highly charged system: besides 4 acidic residues, 10 basic residues are found therein. Such highly complex systems were studied by pK_a calculations for aldose reductase.[25] If possible, ITC measurements were of great help for a deeper understanding of the binding process. Such measurements could assist in deciphering the overall protonation effect of ligand binding.

5.2. FXa

A clear indication for the influence of protein–ligand interactions on protonation states is the *BIPS*/*aBIPS* switch of ASP-189 in the S1 pocket. The difference in the mismatch of the PROPKA and MCCE results can be explained by the treatment of the degree of solvent accessibility. In the uncomplexed form, ASP-189 shows a solvent accessibility of 9% which is reduced to 0% for the complex. In the PROPKA computations, the desolvation contribution is the main driving force and gives a pK_a shift of almost 4 log units only marginally shifted by the site–site interactions. In the MCCE computations, the apo and the holo form differ in the size of the desolvation contribution. In the MCCE-FULL/eps = 4 calculation, the apo form gives a pK_a shift of 5 log units by means of desolvation, whereas the holo form gives a pK_a shift of 9 log units by means of desolvation. In both cases, this contribution is further damped by the site–site interactions to ASP-194 and the A-chain N-terminus. Whether or not a protonation change is occurring for ASP-189 in the S1-pocket remains to be revealed by experiments. In a former study on pK_a effects for trypsin and thrombin complexes, we had access to such experimental data.[26] We did not observe a similar effect for ASP-189 for the closely related serine proteases trypsin and thrombin. Instead, a protonation change was found for HIS-57, a member of the catalytic triad. However, the studied ligand series contained a benzamidine anchor in the S1 pocket which introduces a positive formal charge in the S1 pocket other than the neutral formal charge of the chlor-thiophene anchor in the case of rivaroxaban.

5.3. HIV-RT

None of the active site residues shows a perturbed *BIPS* value. The drug rilpivirine induces no switch of the *BIPS* values in the active site as well. One MCCE calculation (QUICK/eps = 4) gives significantly more *aBIPS* values than all other MCCE calculations. The majority of these residues are lysine residues being largely solvent-exposed. Apparently, the reduced sampling in combination with a low dielectric constant gives an over-estimation of the site–site interactions which are responsible for the *aBIPS* values. Although the remaining MCCE calculations do not give that many *aBIPS* values, the trend of the computed pK_a values continues in the same line (data not shown in Table 3).

5.4. PDE4D

In the case of GLU-339, the *aBIPS* prediction by PROPKA stems from the desolvation term. Since GLU-339 has no contact to the drug roflumilast, its pK_a value is only slightly modulated upon complexation. Thus, GLU-339 also shows an *aBIPS* value in the complexed form. TYR-303 is an example for the sensitivity of the *BIPS* measure: In the DEFAULT-MCCE/eps = 4 calculation, it has a pK_a value of 8.15 (15% de-protonated) in the complexed form versus 8.56 (6% de-protonated) for the apo form. Although the pK_a shift amounts to less than half a log unit, this leads to a *BIPS*/*aBIPS* switch. All other calculations show a clear preponderance of the protonated form of TYR-303, resulting in a typical *BIPS* value. Another example is HIS-105: The QUICK-MCCE/eps = 8 calculation indicates a perturbed *BIPS* value. However, this pK_a value is just at the borderline for a *BIPS* transition, and all other MCCE calculations give a normal *BIPS* value. The maximum deviation between the computed pK_a values of HIS-105 amounts to 0.9 log units. Therefore, it is reasonable to regard the single *aBIPS* value of HIS-105 as outlier. The binding site (6 Å radius around the ligand) is composed of acidic and basic residues in an equilibrated manner: 5 acids (ASP-201, GLU-230, ASP-272, ASP-318 and ASP-334) and 5 bases (TYR-159, HIS-160, HIS-204, HIS-315 and TYR-329). Of

these, only HIS-160 and HIS-315 give rise to *aBIPS* values. Upon complex formation with roflumilast, these residues show pK_a shifts/*aBIPS* transformations pointing into different directions. Thus, we postulate a proton shuffle mechanism, but further modelling would be necessary to prove this hypothesis. Studies on a thioredoxin-like protein[27] and serine carboxyl peptidase[28] were able to prove such proton shuffling mechanisms. However, the applied computational techniques such as MD simulations or QM/MM go beyond the scope of this contribution. PDE4D seems to be a special case with regard to the preponderance of histidines showing *aBIPS* values. It is remarkable that the 'histidine density' is rather high for PDE4D in contrast to the other studied protein–ligand complexes of our study: 5% (HIS count divided by the overall residue number) versus 2%, on average, for the other studied complexes.

5.5. RET

Although the interplay between the *aBIPS* values of PTR-905 and LYS-907 is only of minor relevance for drug design, this titration behavior might be of relevance for the phosphorylation step of RET tyrosine kinase. From a mechanistic point of view, such an effect is exciting but further theoretical investigations are not within the scope of this publication. Nonetheless, the *BIPS*/*aBIPS* switch of the ligand is clearly dependent on the protein dielectric constant of the MCCE calculations. This switch is caused by the different sign of the site–site interactions. In the case of the low dielectric constant (eps = 4), the site–site interactions are responsible for an upward pK_a shift of the ligand titratable group. In contrast with a high dielectric constant (eps = 8), the site–site interactions shift the pK_a value of the ligand in the opposite direction. Furthermore, if we take into account the normal *BIPS* value of the PROPKA calculation, we believe that the *aBIPS* value of the eps = 4 calculations is a false positive.

5.6. pK_a calculations—an industrial perspective

In structure-based drug design, the effect of protonation changes is often neglected, from the theoretical point of view, as well as from the experiments performed. For the five presented drug–protein complexes, we were able to show that protonation changes upon drug binding are supposed to occur. Therefore, if biophysical data for these drugs were available, we would be able to probe our calculations with reality. To the best of our knowledge, such a detailed analysis is usually not performed in industry, due to time constraints and lack of resources. But not only experiments bind many resources, even our in silico studies are tedious, time-consuming and definitely not an one-stop-shop. The good message from our studies is that the two orthogonal methods, PROPKA and MCCE, often show similar trends regarding the *aBIPS* measure. The advantage of PROPKA is the usability and the speed of the actual calculations. Furthermore, a well designed PROPKA VMD GUI has recently become available simplifying the interpretation of the results.[29] In contrast, the user can perform parameter studies when employing MCCE and gets a very detailed output. The advantage of using both methods is the augmentation of the reliability of the predictions. With our presented *BIPS* measure, one can quickly identify 'pK_a hot spots' in the drug–protein complex. Hopefully, our contribution convinces molecular modelers in industry to consider pK_a calculations as additional information resource, for example when evaluating a new target or prioritizing different chemical series.

6. Conclusions

The central question of this contribution was whether structure-based pK_a calculations are relevant for rational drug design or not. For this purpose, we employed two algorithms (MCCE and PROPKA) for the prediction of protonation states. We were able to show interesting results for three out of five studied drug–protein complexes:

- The catalytic dyad GLU-205, GLU-206 in DPP IV forms a coupled system with the nearby residue ASP-663. The drug linagliptin only shows minor influence on this system.
- In the S1 pocket of factor Xa, we propose that ASP-189 becomes protonated upon complexation with rivaroxaban.
- A proton shuffling mechanism is proposed for two histidine residues in the active site when roflumilast is bound to PDE4D.

We also observe that in some cases the *BIPS* measure is too sensitive and gives false positives. Usually, this occurs only for one parameter setting of MCCE. By a comparison to the other calculations with different parameters, one can quickly identify that the *aBIPS* alert is caused just at the cutoff point of a *BIPS*/*aBIPS* switch. Although the magnitude of the desolvation term seems to be differently weighted by MCCE and PROPKA, the two programs are often aligned in terms of the prediction of *aBIPS* values. Regarding the conformational sampling in MCCE (DEFAULT vs QUICK runs), some differences in the results can be observed. Nonetheless, the overall trends are estimated with both settings.

The *aBIPS* value of the ligand titratable group rilpivirine in the MCCE/eps = 8 calculation can be also considered as false positive. Here, the majority of the pK_a calculations (MCCE/eps = 4 and PROPKA) gives a normal *BIPS* value which leads us to this judgement. An even better agreement is found for residues with typical *BIPS* values; we only look rather superficially at these 'pK_a cold spots'. Ultimately, we recommend using both pK_a calculation methods, PROPKA and MCCE, for a profound judgment on atypical protonation states.

Acknowledgments

My sincere gratitude goes to Jun Jun Mao from the City College of New York. As one of the MCCE core developers, he very quickly made a new Mac version available to me and was informative source of MCCE wisdom. From my colleagues at Merck KGaA, I would like to thank Friedrich Rippmann, Gerhard Barnickel and especially Daniel Kuhn for their proof reading and the fruitful discussions. Roger Draheim (Goethe University Frankfurt) is highly appreciated for his suggestions concerning semantics and advice on scientific rationales. For the artwork, Steffen Schellenberger is acknowledged.

References and notes

1. Søndergaard, C. R.; Olsson, M.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 525.
2. Song, Y.; Mao, J.; Gunner, M. R. *J. Comput. Chem.* **2009**, *30*, 2231.
3. Davies, M. N.; Toseland, C. P.; Moss, D. S.; Flower, D. R. *BMC Biochem.* **2006**, *7*, 18.
4. Stanton, C. L.; Houk, K. N. *J. Chem. Theory Comput.* **2008**, *4*, 951.
5. Czodrowski, P. *Proteins: Structure, Function, and Bioinformatics* **2011**, *79*, 3299.
6. Maestro version 9.2, Schrodinger LLC, New York, NY, USA 2011.
7. Eckhardt, M.; Langkopf, E.; Mark, M.; Tadayyon, M.; Thomas, L.; Nar, H.; Pfrengle, W.; Guth, B.; Lotz, R.; Sieger, P.; Fuchs, H.; Himmelsbach, F. *J. Med. Chem.* **2007**, *50*, 6450.
8. Roehrig, S.; Straub, A.; Pohlmann, J.; Lampe, T.; Pernerstorfer, J.; Schlemmer, K.-h.; Reinemer, P.; Perzborn, E. *J. Med. Chem.* **2005**, *48*, 5900–5908.
9. Das, K.; Bauman, J. D.; Clark, A. D.; Frenkel, Y. V.; Lewi, P. J.; Shatkin, A. J.; Hughes, S. H.; Arnold, E. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1466.
10. Card, G. L.; England, B. P.; Suzuki, Y.; Fong, D.; Powell, B.; Lee, B.; Luu, C.; Tabrizizad, M.; Gillette, S.; Ibrahim, P. N.; Artis, D. R.; Bollag, G.; Milburn, M. V.; Kim, S.-H.; Schlessinger, J.; Zhang, K. Y. J. *Structure* **2004**, *12*, 2233.
11. Knowles, P. P.; Murray-Rust, J.; Kjaer, S.; Scott, R. P.; Hanrahan, S.; Santoro, M.; Ibáñez, C. F.; McDonald, N. Q. *J. Biol. Chem.* **2006**, *281*, 33577.
12. PRIME 3.0, Schrodinger LLC, New York, NY, USA 2011.
13. Czodrowski, P.; Dramburg, I.; Sotriffer, C. A.; Klebe, G. *Proteins: Structure, Function, and Bioinformatics* **2006**, *65*, 424.

14. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.
15. pKalc 4.3.0, CompuDrug, Sedona, AZ, USA 2011.
16. Warshel, A. *Biochemistry* **1981**, *20*, 3167–3177.
17. Schutz, C. N.; Warshel, A. *Proteins: Structure, Function, and Bioinformatics* **2001**, *44*, 400.
18. Klingen, A. R.; Bombarda, E.; Ullmann, G. M. *Photochem. Photobiol. Sci.* **2006**, *5*, 588.
19. Søndergaard, C. R.; McIntosh, L. P.; Pollastri, G.; Nielsen, J. E. *J. Mol. Biol.* **2008**, *376*, 269.
20. McIntosh, L. P.; Hand, G.; Johnson, P. E.; Joshi, M. D.; Körner, M.; Plesniak, L. A.; Ziser, L.; Wakarchuk, W. W.; Withers, S. G. *Biochemistry* **1996**, *35*, 9958.
21. Qin, J.; Clore, G. M.; Gronenborn, a. M. *Biochemistry* **1996**, *35*, 7.
22. Webb, H.; Tynan-Connolly, B. M.; Lee, G. M.; Farrell, D.; O'Meara, F.; Søndergaard, C. R.; Teilum, K.; Hewage, C.; McIntosh, L. P.; Nielsen, J. E. *Proteins* **2011**, *79*, 685.
23. Tambuyzer, L.; Azijn, H.; Rimsky, L. T.; Vingerhoets, J.; Lecocq, P.; Kraus, G.; Picchio, G.; de Béthune, M.-P. *Antivir. Ther.* **2009**, *14*, 103.
24. Johnson, L. B.; Saravolatz, L. D. *Clin. Infect. Dis.* **2009**, *48*, 1123.
25. Steuber, H.; Czodrowski, P.; Sotriffer, C. A.; Klebe, G. *J. Mol. Biol.* **2007**, *373*, 1305.
26. Czodrowski, P.; Sotriffer, C. A.; Klebe, G. *J. Mol. Biol.* **2007**, *367*, 1347.
27. Guo, H.; Wlodawer, A.; Nakayama, T.; Xu, Q.; Guo, H. *Biochemistry* **2006**, *45*, 9129.
28. Narzi, D.; Siu, S. W. I.; Stirnimann, C. U.; Grimshaw, J. P. a.; Glockshuber, R.; Capitani, G.; Böckmann, R. a. *J. Mol. Biol.* **2008**, *382*, 978.
29. Rostkowski, M.; Olsson, M. H. M.; Søndergaard, C. R.; Jensen, J. H. *BMC Struct. Biol.* **2011**, *11*, 6.
30. Stierand, K.; Maass, P. C.; Rarey, M. *Bioinformatics (Oxford, England)* **2006**, *22*, 1710 (http://www.poseview.de).