Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2007) 429–433
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

A simple method to calculate the accessible volume of protein-bound ligands: Application for ligand selectivity

Erlend Hodneland, Knut Teigen *

Department of Biomedicine, University of Bergen, 5009 Bergen, Norway

Received 30 October 2006; received in revised form 25 January 2007; accepted 26 January 2007
Available online 3 February 2007

Abstract

We here present a method for estimating the accessible volume surrounding each atom of a ligand when bound to a receptor. The accessible volumes are calculated as the volume of a sphere with fixed radius, centered on each ligand atom, not simultaneously occupied by the volume of the receptor. The method is capable of describing the packing of the ligand in the binding pocket. Moreover, where structural models of several receptors in complex with the same ligand are available, a comparative study discriminating on accessible volumes can be performed. For these cases, a relatively large accessible volume in a particular receptor might indicate that this receptor has a unique cavity that might be exploited to develop a selective ligand analog. The ligand atoms showing variation in surrounding volume accessibility when bound to different receptors constitute attractive anchor points where one might want to attach substituents that modify the selectivity of the ligand. We have applied the described method to two different enzyme-ligand systems that bind tetrahydrobiopterin, i.e. the aromatic amino acid hydroxylases and nitric oxygen synthases. Our results yield new insights into the specificity of cofactor binding to these protein families.

Keywords: Binding site analysis; Molecular shape; Protein–ligand interactions; Lead optimization; Drug design; Volume estimations

1. Introduction

An essential prerequisite for structure based drug design is a comprehensive structural understanding of the drug target at hand. Binding pockets or cavities within proteins are often attractive as possible interaction faces for a new drug. Several computer algorithms have been developed to screen protein structures to find invaginations of the surface that might be used as a starting point for the development of a compound for specific pharmacological regulation. Other computer algorithms seek to perform a qualitative analysis of the binding pocket at hand, usually by applying a force-field based approach [1–7].

We initiated the work described in this study when we studied selectivity determinants for cofactor binding to the aromatic amino acid hydroxylases (AAH) [8]. It proved difficult to get an impression of the *difference* in available volume around the cofactor when bound to each of the three binding pockets of the AAHs. We were not able to find any software that was able to compute and visualize this satisfactorily. Thus, we embarked on developing the herein described algorithm that provides us with an intuitive representation of the accessible volume surrounding a ligand when bound to a receptor. The shape complementarity is a required but not a sufficient criterion for molecular recognition. The proposed shape based parameter could be utilized in lead optimization in combination with other molecular interaction parameters. Although the shape complementarity per se is easily observable by visual inspection, its quantitative expression can be used to get a measure of the difference in accessibility between receptors interacting with the same ligand. The algorithm is applied to investigate the difference in volume around the same ligand when bound to proteins from two enzyme families.

Abbreviations: AAH, aromatic amino acid hydroxylases; BH2, L-*erythro*-7,8-dihydrobiopterin; BH4, (6*R*)-L-*erythro*-5,6,7,8-tetrahydrobiopterin; eNOS, endothelial nitric oxide synthase; H_4-pyrazine, tetrahydro-pyrazine; iNOS, inducible nitric oxide synthase; NOS, nitric oxide synthase; PAH, phenylalanine hydroxylase; nNOS, neuronal nitric oxide synthase; TH, tyrosine hydroxylase; TPH, tryptophan hydroxylase

* Corresponding author. Tel.: +47 55586328; fax: +47 55586360.
E-mail address: knut.teigen@biomed.uib.no (K. Teigen).

1093-3263/$ – see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2007.01.011

2. Methods

2.1. Volume estimations

The requirement of this method is that a structural model of a ligand in complex with a receptor is available, either from experimental techniques or from theoretical modeling. The ligand is initially removed from the complex, and the open-source program DMS (http://www.cgl.ucsf.edu/Overview/software.html#dms) is used to describe the molecular surface of the ligand binding pocket. A united radius of each receptor atom is defined based on the atom type. The surface resembles the van der Waals surface of the receptor, except that crevices between atoms are smoothed over and interstices too small to accommodate a probe with default radius of 1.4 Å (resembling a water molecule) are eliminated. The surface calculated is that defined by Richards [9] and referred to as the re-entrant molecular surface. The density of points generated is approximately 20 points per square Å. The structure of the receptor is then discarded, and the re-entrant molecular surface together with the ligand atoms is used in further analysis. To obtain a volume estimation, the Euclidian distances between each atom and the receptor surface are calculated for a set of discretized directions surrounding the ligand atom. To accomplish this, let $\vec{s}_{\varphi,\theta}$ be a unit vector pointing in the direction specified by the spherical coordinates $\theta \in \langle 0, \pi]$, $\varphi \in \langle 0, 2\pi]$ where θ is the angle to the positive x-axis and φ is the angle to the positive z-axis (Fig. 1).

Starting at the ligand atom and performing incremental steps $\Delta s_{\varphi,\theta}$ along $\vec{s}_{\varphi,\theta}$ (Fig. 2, top) until cut-off value $C = N\Delta s_{\varphi,\theta}$, the shortest distance $d(r)$ to the receptor surface is calculated for each step where $r = n\Delta s_{\varphi,\theta}$ is the distance from the ligand atom to the current position along $\vec{s}_{\varphi,\theta}$, $n = \{1, \ldots, N\}$. Thus, a distance profile $d(r)$ along $\vec{s}_{\varphi,\theta}$ is obtained, at each element containing the shortest distance to the receptor surface (Fig. 2, bottom). The desired distance between the ligand atom and the receptor surface is then obtained by calculating the argument providing the smallest local minimum value of $d(r)$ which is smaller than the maximum distance between surface points δ. Inherently, this formulation will provide the distance from the ligand atom to the first receptor surface that is encountered in the incremental step procedure. In the cases when $d(r)$ at no point drops below δ, a receptor surface has not been found in this specific direction and the cut-off value itself is designated as the desired distance. These formulations can be described mathematically as picking out the minimum value of the set of local minima of $d(r)$, $R(\varphi, \theta) = \min\{\arg\min_r(d(r))\}$ where $\arg\min_r(d(r))$ here returns the set of all r that minimize $d(r)$ below δ. If no surface is found, corresponding to $d(r) > \delta\ \forall r$, R is assigned the cut-off value itself $R(\varphi, \theta) = C$.

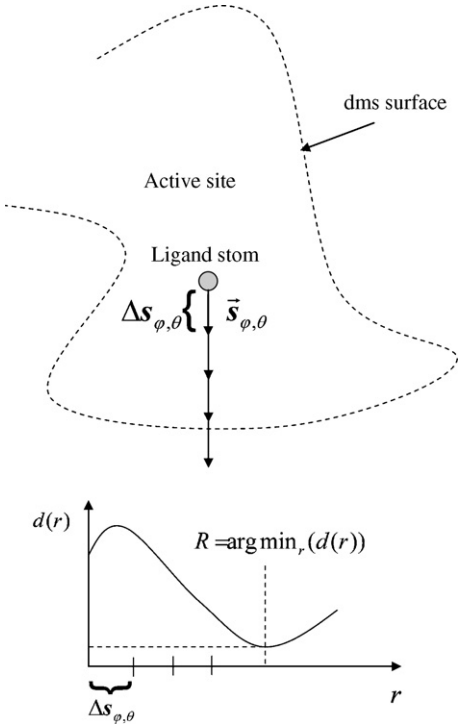

Fig. 2. Estimating the distance from a ligand atom to the receptor surface. For each step $r = n\Delta s_{\varphi,\theta}$ in direction of $\vec{s}_{\varphi,\theta}$ ($n = \{1, \ldots, N\}$ where $N\Delta s_{\varphi,\theta}$ equals the pre-set cut-off value), the shortest distance $d(r)$ to the receptor surface is calculated. The obtained distance profile $d(r)$ is used to compute the distance from the ligand to the molecular surface which appears as the argument minimizing $d(r)$. If the re-entrant surface is to a great extent curved, several surfaces are detected in this process. Therefore, the final distance from the ligand atom to the receptor surface is calculated as the minimum within R.

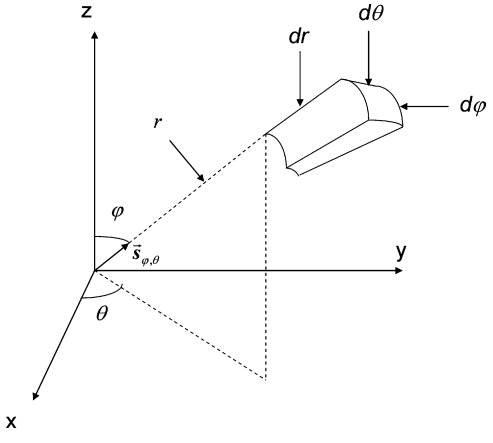

Fig. 1. The coordinate transformation from Cartesian to spherical used for volume integration. The volume integration is applied to a small volume element specified by the incremental distance dr, the incremental angle dθ to the positive x-axis and the incremental angle dφ to the positive z-axis. The distance from the ligand atom to the small volume element along the direction of $\vec{s}_{\varphi,\theta}$ is given by r.

The obtained distance-measures are used for integration to calculate the desired volume. Let the discrete set of distances $R(\varphi, \theta)$ span the volume of interest, then the volume v is given by the following equation:

$$v = \iiint_V \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = \int_0^{2\pi}\int_0^{\pi}\int_0^{R(\varphi,\theta)} r^2 \sin\phi\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\phi \qquad (1)$$

after a transformation from Cartesian to spherical coordinates [10] (Fig. 1).

For the numerical analysis, a discretized version of Eq. (1) is being used,

$$v = \sum_{0}^{2\pi}\sum_{0}^{\pi}\sum_{0}^{R(\phi,\theta)} r^2 \sin\phi\, \Delta r\, \Delta\theta\, \Delta\phi.$$

The discretizing steps used in the calculations are $\Delta r = 0.1$ Å, $\Delta\phi = \Delta\theta = 0.1$ rad. To restrict the computational time, all surface points outside the maximum cut-off distance were removed prior to calculations. Furthermore, between each subsequent step $\Delta \boldsymbol{s}_{\varphi,\theta}$ only the surface points that approached, were included in the next step.

As an alternative approach producing approximately equal results, it is possible to define a set of N evenly spaced points at a certain distance around each ligand atom. The available volume is then calculated as a sum of the differential volumes belonging to each point. This formulation could save some computational time as we avoid uneven spacing between search directions, but on the other hand it requires an additional algorithm to find an even distribution of points.

2.2. Test systems

To test the method described in this paper, we have applied it to two distinct enzyme systems that both bind the same ligand, L-*erythro*-5,6,7,8-tetrahydrobiopterin (BH4, see Fig. 3). This cofactor is shared between the members of the aromatic amino acid hydroxylase (AAH) family; phenylalanine hydroxylase (PAH, EC 1.14.16.1), tyrosine hydroxylase (TH, EC 1.14.16.2) and tryptophan hydroxylase (TPH, EC 1.14.16.4) and the three different isoforms of nitric oxide synthase (NOS, EC 1.14.13.39); endothelial (eNOS), neuronal (nNOS) and inducible (iNOS). Structural models of all these enzymes in complex with BH4 are available. The crystal structure of the human isoforms of eNOS-BH4 and iNOS-BH4 complexes has been solved [11] (pdb accession code 3NOS and 4NOS, respectively) as well as the rat isoform of the nNOS-BH4 complex [12] (pdb accession code 1ZVL). For the hydroxylases, only the crystal structure of PAH in complex with BH4 has been solved [13] (pdb accession code 1J8U). Nevertheless, the crystal structures of human TH [14] and human TPH (isoform 1) [15] in complex with the reduced form of the cofactor, L-*erythro*-7,8-dihydrobiopterin (BH2) has been solved. However, in the model from the crystal structure of the TH-BH2 complex the cofactor is flipped 180° around the C4a–C8a bond with respect to all the other structures of the AAHs

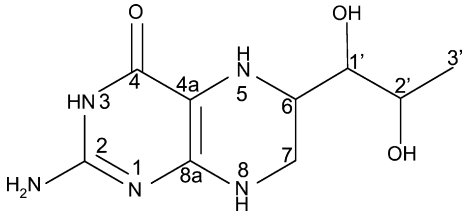

Fig. 3. Chemical structure of the cofactor BH4. The cofactor consists of a fused pyrimidone and tetrahydro-pyrazine ring (shown to the left and right in this figure, respectively). The dihydroxypropyl side chain attached to position 6 is characteristic of *bio*pterin and distinguishes it from other pterin analogs.

solved in complex with the biopterin. Thus, this structural model is most probably a result of misinterpretation of the electron density maps and we have consequently used the structural model of the TH-BH4 and TPH-BH4 complexes as reported in an NMR and molecular modeling study [8]. In order to prepare a model of the TPH-BH4 complex, we reduced the cofactor in the TPH-BH2 complex *in silico* and ran a 1 ns molecular dynamics simulation to relax the structure. The PAH-BH4 and TH-BH4 complexes were also simulated for 1 ns in order to treat all the AAH-BH4 complexes equally. The minimized average structure over the last 100 ps of the simulations was used in the volume calculations. Molecular dynamics simulations were performed with Amber8 as earlier described [16]. The substrate in eNOS and nNOS and substrate analog in iNOS were removed together with water molecules from all the models before the accessible volume around BH4 was calculated.

3. Results

We used the herein described method to calculate the accessible volume in the immediate proximity of tetrahydrobiopterin (BH4) when bound to six different enzymes. To get an impression of the packing of BH4 in the NOS binding pocket, the accessible volumes for the individual atoms of BH4 when bound to the NOS isoforms are shown in Fig. 4A. There are two distinct BH4 binding sites in the interface between the two monomers in a NOS dimer. In addition, for iNOS it was reported two biologically relevant dimers per asymmetric unit [11], thus adding up to a total of four models of the BH4 binding site in iNOS. These were all treated separately in the analysis. However, the difference between the binding sites within each NOS isoform was found to be negligible compared to the accessible volume difference between isoforms (data not shown). Thus, the representation in Fig. 4A shows the averaged accessible volumes around BH4 when bound to each of the different NOS isoforms. The same algorithm was then used to calculate the accessible volume around each atom of BH4 while bound to the cofactor binding site of the three members of the AAH family. The results are represented graphically in Fig. 4B. The maximum threshold value used for the incremental search protruding from the ligand atom to calculate the accessible volume was set to 5 Å. Thus, the maximum accessible volume that an atom can obtain is 523.6 $Å^3$. A larger maximum threshold value might explore the surface farther away from each atom, but the value obtained will be less unique to the atom it is calculated for. In our experience, a maximum threshold value of 5 Å gives a good compromise between searching a large enough volume while still obtaining local volume information around each atom. To get a better impression of the difference in the accessible volume between the two enzyme systems, the accessible volumes of BH4 atoms in TPH and eNOS are shown together in Fig. 4C.

4. Discussion

The accessible volumes are fast to compute and illustrate very well the available volume around each atom when bound

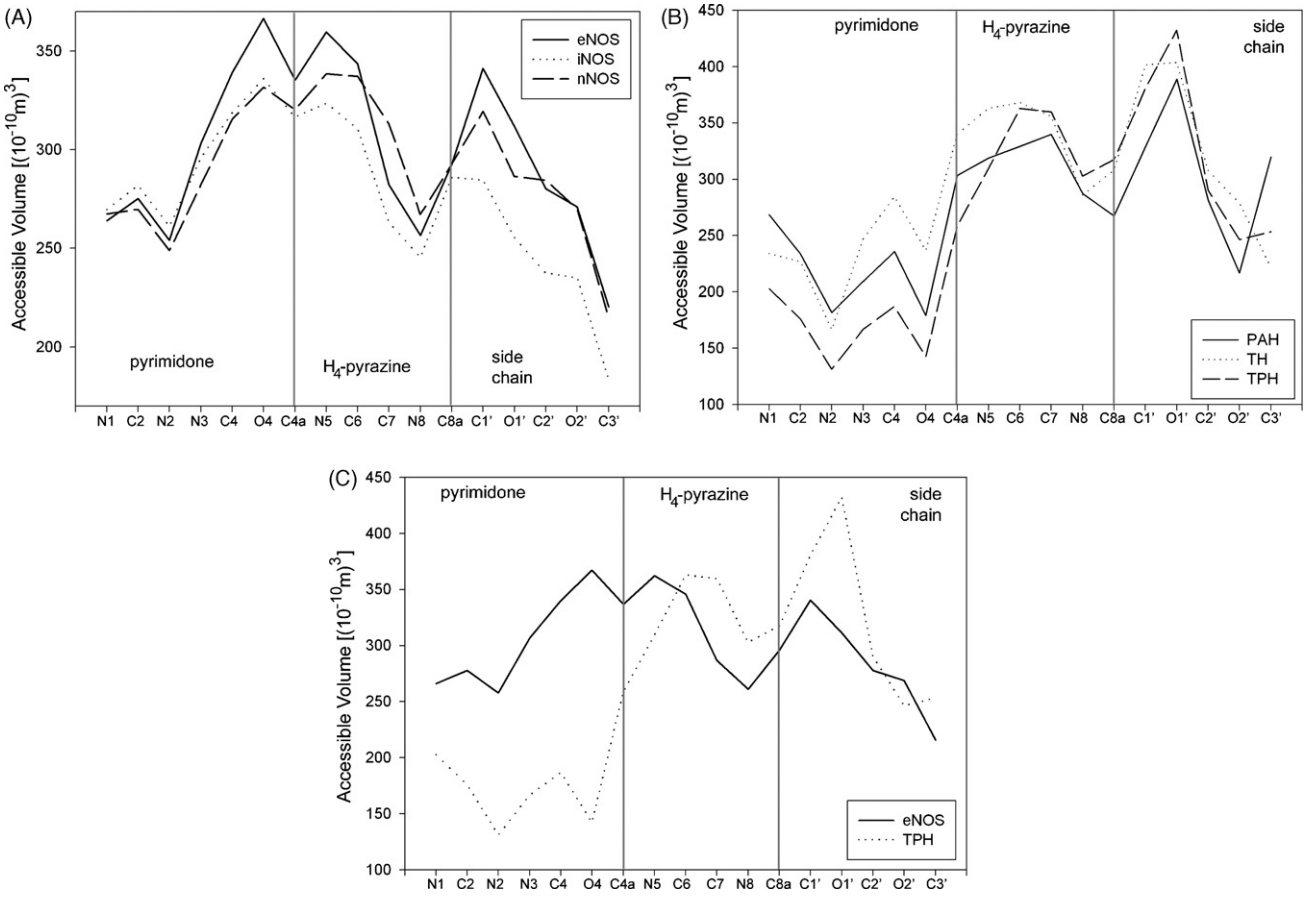

Fig. 4. Accessible volume around the atoms of BH4 while bound to; (A) eNOS, iNOS and nNOS, (B) PAH, TH and TPH and (C) TPH and eNOS. The labeling along the *x*-axis corresponds to the atom numbering of BH4 in Fig. 3.

to the protein pocket. As a general trend, it appears to be a difference in accessible volume on each side of the ring systems in BH4 when bound to NOS (Fig. 4A). The atoms that constitute the lower part in Fig. 3 (N2, C2, N1, C8A and C8), have a smaller volume accessible as compared to the atoms on the opposite side (C4, O4, C4A, N5 and N6, together with C1′) indicating a higher tolerance of cofactor analogs with bulky substituents in these positions. The accessibility of BH4 is plotted for all the three NOS isoforms in Fig. 4A so that we can do a comparative study of the accessibility of the BH4 atoms when bound to the NOS isoforms. From the plot we see that the accessible volume around the BH4 dihydroxypropyl side chain is larger for nNOS and eNOS as compared to iNOS. This corresponds well with the finding that BH4 analogs with bulky substituents at C6 are not very well tolerated by iNOS, while eNOS and nNOS can accommodate these analogs better, as discussed in the review by Matter et al. [12]. It seems to be a general trend that nNOS is able to accommodate larger substituents [12,17] as compared to both iNOS and eNOS. However, our analysis indicates that eNOS has the largest binding pocket for BH4 of the three NOS isoforms. It is important to keep in mind that the structure used in this study for nNOS is from rat, while the eNOS and iNOS isoforms are from human. Thus, the accessible volumes in rat nNOS might not be relevant in a discussion of comparative studies of BH4 analogs with specificity for selected human isoforms of NOS.

All the models of the hydroxylases in complex with BH4 used in this study are with the human form of the enzymes. As seen in Fig. 4B, they all show comparable accessible volume with minor variations for accommodating BH4 in the active site. It has already been pointed out that analogs of BH4 with substituents at N8 might serve as a scaffold for TPH selective cofactor analogs [8]. In agreement with this, our analysis shows that the volume accessible to N8 is slightly larger in TPH as compared to the other hydroxylases.

The size of the BH4 binding pocket is very similar within the NOS isoforms and within the members of the AAH family. However, if we compare the binding site of one arbitrary representative from each family, there are some striking differences (Fig. 4C). The accessible volume around the pyrimidone ring is significantly larger for the NOS family as compared to the AAHs, while the hydroxylases seem to have more available volume surrounding the side chain. Thus, it seems like an attractive approach to add substituents to the pyrimidone ring to make cofactor analogs selective for the NOS family, while bulky side chains at the 6-position of the tetrahydro-pyrazine ring might be better tolerated by the AAH family.

5. Conclusions

The algorithm described in this paper is fast to compute and give a very intuitive description of the accessible volume

surrounding a ligand in the binding pocket. This method can be used comparatively to study the difference in accessible volume of the same ligand when bound to different receptors. The information obtained from such an analysis can be used to determine which receptors might better accommodate substituents at specific atoms of the ligand or lead compound.

Acknowledgements

Aurora Martínez is thanked for valuable discussions and for carefully reading the manuscript. This work was supported by grants from The Research Council of Norway and The Norwegian Cancer Society (grant number A05103/004).

References

[1] C.A. Del Carpio, Y. Takahashi, S. Sasaki, A new approach to the automatic identification of candidates for ligand receptor sites in proteins. (I). Search for pocket regions, J. Mol. Graph. 11 (1) (1993) 23–29, 42.
[2] D.R. Flower, Receptor-binding sites: bioinformatic approaches, Methods Mol. Biol. 316 (2006) 291–358.
[3] P.J. Goodford, A computational procedure for determining energetically favorable binding sites on biologically important macromolecules, J. Med. Chem. 28 (7) (1985) 849–857.
[4] M. Hendlich, F. Rippmann, G. Barnickel, LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins, J. Mol. Graph. Model. 15 (6) (1997) 359–363, 389.
[5] J. Liang, H. Edelsbrunner, P. Fu, P.V. Sudhakar, S. Subramaniam, Analytical shape computation of macromolecules. I. Molecular area and volume through alpha shape, Proteins 33 (1) (1998) 1–17.
[6] J. Liang, H. Edelsbrunner, C. Woodward, Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design, Protein Sci. 7 (9) (1998) 1884–1897.
[7] A. Miranker, M. Karplus, Functionality maps of binding sites: a multiple copy simultaneous search method, Proteins 11 (1) (1991) 29–34.
[8] K. Teigen, K.K. Dao, J.A. McKinney, A.C. Gorren, B. Mayer, N.A. Frøystein, A. Martínez, Tetrahydrobiopterin binding to aromatic amino acid hydroxylases. Ligand recognition and specificity, J. Med. Chem. 47 (24) (2004) 5962–5971.
[9] F.M. Richards, Areas, volumes, packing and protein structure, Annu. Rev. Biophys. Bioeng. 6 (1977) 151–176.
[10] T. Apostol, Calculus, John Wiley & Sons, Inc., United States of America, 1969.
[11] T.O. Fischmann, A. Hruza, X.D. Niu, J.D. Fossetta, C.A. Lunn, E. Dolphin, A.J. Prongay, P. Reichert, D.J. Lundell, S.K. Narula, P.C. Weber, Structural characterization of nitric oxide synthase isoforms reveals striking active-site conservation, Nat. Struct. Biol. 6 (3) (1999) 233–242.
[12] H. Matter, H.S. Kumar, R. Fedorov, A. Frey, P. Kotsonis, E. Hartmann, A.J. Prongay, P. Reichert, D.J. Lundell, S.K. Narula, P.C. Weber, Structural analysis of isoform-specific inhibitors targeting the tetrahydrobiopterin binding site of human nitric oxide synthases, J. Med. Chem. 48 (15) (2005) 4783–4792.
[13] O.A. Andersen, T. Flatmark, E. Hough, High resolution crystal structures of the catalytic domain of human phenylalanine hydroxylase in its catalytically active Fe(II) form and binary complex with tetrahydrobiopterin, J. Mol. Biol. 314 (2) (2001) 279–291.
[14] K.E. Goodwill, C. Sabatier, C. Marks, R. Raag, P.F. Fitzpatrick, R.C. Stevens, Crystal structure of tyrosine hydroxylase at 2.3 A and its implications for inherited neurodegenerative diseases, Nat. Struct. Biol. 4 (7) (1997) 578–585.
[15] L. Wang, H. Erlandsen, J. Haavik, P.M. Knappskog, R.C. Stevens, Three-dimensional structure of human tryptophan hydroxylase and its implications for the biosynthesis of the neurotransmitters serotonin and melatonin, Biochemistry 41 (42) (2002) 12569–12574.
[16] K.K. Dao, K. Teigen, R. Kopperud, E. Hodneland, F. Schwede, A.E. Christensen, A. Martínez, S.O. Døskeland, Epac1 and cAMP-dependent protein kinase holoenzyme have similar cAMP affinity, but their cAMP domains have distinct structural features and cyclic nucleotide recognition, J. Biol. Chem. 281 (30) (2006) 21500–21511.
[17] H. Matter, P. Kotsonis, Biology and chemistry of the inhibition of nitric oxide synthases by pteridine-derivatives as therapeutic agents, Med. Res. Rev. 24 (5) (2004) 662–684.