**Dynamics fingerprints of active conformers of epidermal growth factor receptor kinase.**

G. P. Barletta, M. Hasenhuer, M. S. Fornasari, G. Parisi, and S. Fernandez-Alberti\*

Universidad Nacional de Quilmes/CONICET, Roque Saenz Peña 352, B1876BXD Bernal, Argentina

ABSTRACT

Epidermal growth factor receptor (EGFR) is a prototypical cell-surface receptor that plays a key role in the regulation of cellular signaling, proliferation and differentiation. Mutations of its kinase domain have been associated with the development of a variety of cancers and, therefore, it has been the target of drug design. Single amino acid substitutions (SASs) in this domain have been proven to alter the equilibrium of pre-existing conformer populations. Despite the advances in structural descriptions of its so-called active and inactive conformations, the associated dynamics aspects that characterize them have not been thoroughly studied yet. Using normal mode analysis, we have extensively explored global dynamics aspects that differentiate these conformers. For this purpose, we present a novel procedure to identify the collective dynamics shared among different conformers in EGFR kinase. The method allows the comparison of patterns of low-frequency vibrational modes defining representative directions of motions. Our procedure is able to emphasize the main similarities and differences between effects that mutations have on the collective dynamics of the protein. In the case of EGFR kinase, two representative directions of motions have been found as dynamics fingerprints of the active conformers. Otherwise, the inactive conformers exhibit a more heterogeneous distribution of collective motions.

## I. INTRODUCTION

The epidermal growth factor receptor (EGFR) is a key protein in cellular signaling, and regulation of cell proliferation, differentiation, and migration1. Mutations and overexpression of EGFR have been associated to different human cancers2,3 and its kinase domain (EGFR kinase) is an intensely pursued target of small-molecule drugs4,5,6,7. Single amino acid substitutions (SASs), deletions and insertions in this domain can alter the equilibrium of pre-existing conformer populations89,10. In this way, certain EGFR kinase cancer mutants become enzymatically more active than the wild-type11,12. These SAS are denominated as activating, due to the stabilization of the conformer required to drive the phosphorylation (active form).

The detailed understanding at a molecular level of the structural and dynamical differentiation among wild-type and cancer mutants of EGFR kinase is a very important goal in the development of personalized cancer treatments13. In particular, it would be most enlightening to be able to elucidate structural and dynamical features that represent fingerprints of kinase activation.

The large variety of available experimental structures reveals the presence of at least two main EGFR conformations, the so-called active and inactive conformers14,15. Furthermore, common sequence and structural characteristics shared among most of active EGFR kinase mutants have been identified16,17 (see Figure 1): (a) α C-helix rotated inwards against the N-lobe and towards the active site, allowing a salt-bridge between residues E762 and K745 (according to the human EGFR canonical amino acid sequence numbering (Universal Protein Resource, UniProtKB, P00533, isoform 1)) 16; (b) a more extended and open conformation of the activation loop (residues 855-886), so-called DFG-in conformation with its aspartate pointing to the ATP-binding site. Besides these two main features, the triad HRD (residues 835-837) 18,19 and different amino acid networks seem to contribute to the ATP active site stabilization 20,21,22. Nevertheless, not all of active conformers fulfill all of this requirements23, and several reported controversies and ambiguous conformation classification can be found24. In a previous article25, we have extensively analyzed these structural parameters in a large set of human EGFR kinase domains previously classified as active or inactive forms. We have focused our attention on changes in the size and shape of the main pocket among the EGFR kinase active and inactive conformers. For this purpose, a hierarchical clustering based on RMSD (root mean square difference) of α-C belonging to main pocket positions has been performed. Our main pocket structural comparison allows us to analyze effects that changes in the main pockets’ structural features have on the EGFR kinase activity.

Despite advances in structural descriptions of the so-called active and inactive conformations26,27,28,17,29, the associated dynamics aspects that characterize them have not been thoroughly studied yet. Using unbiased long-timescale molecular dynamics simulations, Yan et al. 30 have explored the transitions from the active to the inactive conformations of a single EGFR mutant. Besides, the intrinsic disorder of EGFR kinase domain has been shown to play a significant role in the EGFR dimerization31.

In the present study, we use normal mode analysis (NMA) to explore global dynamics aspects that differentiate these conformers. NMA based on a coarse-grained model of proteins has been extensively proved as a useful technique to explore the intrinsic dynamics within a folded state32,33,34,35,36. The complex motions of proteins are decoupled into a linear combination of independent harmonic oscillators, i.e., the normal modes. Low-frequency normal modes correspond to the most collective ones, involving the concerted motions of many atoms. These modes are commonly related to functional aspects of a protein37,38,39,40 and they have been proven to be robust to sequence variations, that is, they are evolutionary conserved41,42,43,44,45.

The fact that normal modes provide a decoupled harmonic description of protein vibrations is fundamental to identify common dynamics aspect shared by a set of related proteins44. Nevertheless, small local structural perturbations, introduced by SASs, can significantly modify the global dynamics and functionality of proteins46.

Herein, we present a novel procedure that allows us to identify collective dynamics shared among different conformers in EGFR kinase. The method allows the comparison of patterns of low-frequency vibrational modes defining representative directions of motions that can be considered as dynamics fingerprints of the active conformers.

The paper is organized as follows. Theoretical methods are described in Section II. Our results are presented and discussed in Section III. Finally, Section IV summarizes our findings and conclusions.

**II. METHODS**

1. *Set of active and inactive EGFR kinase structures*

Our set of active and inactive EGFR kinase structures has been selected from CoDNAS47 corresponding to the canonical sequence of the human EGFR (UniProtKB id: P00533). EGFR kinase is a domain with *N*=277 residues plus the activation loop. Structures presenting more than 16 missing residues were not included in our dataset, otherwise missing residues have been included using the MODELLER software package with loop optimization 48. Most of the missing residues were located in the activation loop and a loop between the β-sheets of the N lobe. Each model was evaluated using the DOPE scoring function to assert that each amino acid scored a negative energy value. Theoretical B-factors profiles have been calculated using NMA on each protein of the initial dataset and compared them with the experimental B-factor profiles. Then, we removed from the data set all structures with optimal Spearman rank correlation coefficient between experimental and theoretical B-factors < 0.6 Å. As a result of these filters in the original dataset, we were left with a final dataset composed of 56(**nuevamente chequear la compatibilidad con el aticulo de Marcia, no deberiamos tener ninguna que Marcia haya descartado**) EGFR kinase structures, 36 of which are classified as active conformers and the rest of 20 structures as inactive conformers. The complete set of EGFR kinase structures used in our final data set is given in **Supplementary Info files**.

1. *NMA background*

Normal mode analysis have been performed using the Elastic Network Model (ENM) that considers the protein as an elastic network of *N* -carbons (nodes) linked by springs within a cutoff distance *rc*. For each protein in the dataset, the value of *rc* varies in order to obtain the best match between theoretical and experimental B-factors.

The interaction potential between residues is defined as32,49, 50

(1)

being the vector connecting residues *i* and *j*, and the zero superscript indicates the equilibrium position that corresponds to the coordinates of the -carbons in the experimental structure. The value of the force constant is determined according to the type of interaction between nodes51:



else

if then

if *i* and *j* are connected by disulphide bridge 

if *i* and *j* interact by hydrogen bond or salt bridge  x 0.1

otherwise  x 0.01

if 

being ** a scaling constant to match the theoretical result to experimental data. CSU program52 has been used to obtain the connectivity information related to hydrogen bonds, salt bridges, and disulphide bridges.

Normal modes are obtained by solving the eigenvalue equation

(2)

being the Hessian matrix **H** a 3*N*x3*N* matrix of second-order partial derivatives of the potential energy, **q** is an orthogonal *N*x*N* matrix whose columns **q**k are the eigenvectors of **H**, that is, the normal modes, and **** is the diagonal matrix of eigenvalues λk of **H**. The theoretical B-factor *Bi* of residue *i*, associated to its thermal fluctuation, is calculated as53

(3)

with as the mean square displacement from its equilibrium position. that can be expressed as the sum of contributions from the 3N-6 internal modes of motion as54

(4)

where *kB* is the Boltzmann constant, *T* is the absolute temperature.

1. *Comparison of weighted normal mode spaces*

Vibrational motions associated to structural fluctuations on the main pocket of the active site of EGFR kinase structures have been compared through the calculation of the corresponding weighted Gramian matrix55,56,46. Being **q** and the eigenvector 3*N*x3*N* matrices (see Section 2.B) associated to two conformers **A** and **B** respectively, the vector projection of each onto the set of modes is defined as

(5)

where the weight *w*k associated to normal mode is defined as the normalized accumulation of contributions of to B-factors *Bi* of each *i*th residue belongs to the main pocket of the active site of the EGFR kinase

(6)

The weighted Gramian matrix **G** (*N x N*) of the set of vectors is calculated as the matrix of inner products with elements

(7)

The diagonalization of **G**

(8)

allows us to use the eigenvalues of **G**, , as a measure of the similarity between the two vibrational motions. Eigenvalues of **G** varies between 0 and 155. The higher the value of , the more the associated direction of motion, given by the corresponding , is common to both strctures. Therefore, we define the similarity between the two weighted vibrational spaces can be defined as

(9)

Moreover, an effective number of dimensions shared by two structures can be expressed as the Kirkpatrick index 57 rounded to the nearest higher integer

(10)

1. *SVD representative vectors*

The similarity among directions of fluctuations shared by different pairs of structures can be analyzed as follows. Matrices **A***k* of dimension *N x L* are built with columns representing the directions of each of the *L* pairs of structures compared as described in **Sec. 2C**.

(, pair #1) (, pair #2) …. (, pair #*L*) (11)

Singular Value Decomposition (SVD)58 of each **A***k* matrix is performed. That is, each **A***k* is written as the product of an *N x L* column-orthogonal matrix **U***k*, an *L x L*  diagonal matrix **W***k* with positive or zero elements (the singular values), and the transpose of an *L x L*  orthogonal matrix **V***k*:

(12)

Thus, the elements of matrix **A***k* can be expressed as the sum of products of columns of **U***k* and rows of (**V***k*)*T*, with the “weighting factors” being the singular values .

(13)

Because of this, in the present work, the vector with the highest is considered the *representative mode* for the *directions* of the matrix **A***k*.

## III. RESULTS AND DISCUSSION

The human EGFR kinase domain is depicted in **Figure 1**. Following the definition given by Hasenahuer et al.25 using structural an biological information, the main pocket of the active site is selected by manually inspection of the active conformers PDBids 1M14 (apo form) and 2GS6, and considering residues within a 5Å radius from each atom of the ATP analog substrate-peptide conjugate in 2GS6 [(X. Zhang et al. 2006)](https://paperpile.com/c/3Op4ri/IdrIp). Residues belonged to the pocket related to the active site are pointed out in **Figure 2**. The total number of residues in the pocket, *Npocket*, is 65 (**en la ultima version del articulo de Marcia son 53…tendriamos que usar ese numero confirmando que los resultados son los mismos**). Active and inactive conformers (**Figure 1**) present distinguishable structural features concerning relative orientation of αC-helix, the N-lobe and the active site, and changes in the main conformation of the activation loop. Nevertheless, the classification is long from been definitive since not all the currently available structures completely fulfill these requirements.

In the present work, common dynamical features among active conformers are explored in order to achieve unique fingerprints that enlighten us on minima functional mechanisms within them.

We firstly analyze differences in contributions of protein vibrations to thermal fluctuations of main pocket residues. **Figure 3** shows the distributions of relative values of the weight *w*k/*w*1 (see **eq. 6**) associated to each normal mode . The lowest frequency normal mode is the one that contributes the most to thermal fluctuations of main pocket residues. Therefore, *w*k/*w*1 represents the corresponding relative contribution of *k*th normal mode. Small values of *w*k/*w*1 indicate that the contribution of the *k*th normal mode to thermal fluctuations of main pocket residues can be neglected compared to the corresponding contribution of the 1st mode. **Figure 3** shows that, in the case of active conformers, thermal fluctuations of main pocket residues are mainly covered by the first two lowest frequency normal modes. The contribution of the 3rd normal mode results  20% of the contribution of the 1st mode. On the contrary, in the case of inactive conformers, at least the five lowest frequency modes present significant contributions to pocket residue fluctuations. That is, fluctuations of the main pocket of active conformers are restrained to less number of low-frequency collective motions than inactive conformers.

Secondly, we investigate differences in the nature of low frequency normal modes. In order to do that, we have calculated the residue participation number of lowest frequency normal modes for EGFR kinase active and inactive conformers defined by

(14)

where , and are the components of the *i*th atom in the *k*th normal mode. Values of correspond to vibrations equally distributed throughout all the residues of the protein, and corresponds to vibrations involving the displacement of a single residue. In **Figure 4(a)** and **(b)**, we show the distribution of the fraction of residues involved in the motion of the two lowest frequency normal modes, calculated as values *Pk/N*, for active and inactive EGFR kinase conformers respectively. We note first that, on average, mutations that stabilize inactive conformers have a disparity of effects on the collective motion of residues involved in these vibrations, spreading the distributions over a larger range of values. On average, 1st normal modes of active conformers are more delocalized than 1st normal modes of inactive conformers , involving almost the double of residues. The effects on 2nd normal modes are more scattered ( and ) without showing a clear tendency to disrupt the collectivity. That is, indicates a large impact of non-activating mutations on the original collectivity of the lowest normal mode observed on active conformations. It is therefore expected that functional aspects of EGFR kinase involve coordinated motions between residues that are mainly reflected in the lowest frequency normal mode of active conformations.

Normal mode conservation has been shown to increase linearly with collectivity, so that the slowest most collective modes are the most conserved ones45. Therefore, our findings could point out towards a relative conservation of unique flexibility patterns among active conformers compared to inactive ones. In order to analyze this feature, we have calculated Pearson linear correlation *B* coefficients to quantify the similarity between Cα B-factor profiles of pairs of active and inactive conformers. In **Figure 5** we compare the *B* distributions obtained for active and inactive conformers. Means values of 0.92 and 0.82 are obtained for active and inactive conformers respectively. The difference between both distributions is statistically validated by the Kolmogorov-Smirnov (KS) statistic value of 0.57 with a p-value= 1.6x10-22. This result indicates that thermal fluctuations are more conserved between active conformers than between inactive conformers. Nevertheless, B-factors flexibility patterns do not result accurate enough to separate both types of conformers.

Vibrational motions associated to structural fluctuations of pairs of EGFR kinase conformers can be further compared through the calculation of the corresponding weighted Gramian **G** matrix (see **section II.C**). The diagonalization of **G** provides a set of new directions of motion , given in decreasing order of their corresponding eigenvalues . The values of varies in a [0:1] range. On one hand, a value of indicates that the direction of motion defined by the corresponding vector is shared by both conformers. On the other hand, a value of indicates that the corresponding direction is specific to a single conformer. The value of , defined as the average values (see **eq. 9**), provides our final similarity measure between weighted vibrational spaces of **A** and **B** conformers. **Figure 6(a)** displays the distribution of values of obtained over all pairs of active and inactive conformations. The difference between both distributions is statistically validated by the KS statistic value of 0.97 with a p-value= 5.9x10-129. This means that separates active and inactive distributions better than Pearson linear correlation *B* coefficients between α-carbon B-factor profiles. That is, dynamics information concerning the directionality of collective motions allows a significant improvement in the distinction between active and inactive conformations respect to previous results using flexibility patterns (see **Figure 5**).

Furthermore, the effective number of dimensions shared by pairs of conformers, given by the Kirkpatrick index (eq. 10), also allows a clear differentiation between active and inactive conformers. **Figure 6(b)** shows the distributions of values of for both active and inactive conformers. An average value of 1.7 between pairs of active conformers indicates that dynamics similarities between them can be efficiently reduced to 1-2 common directions of motions. On the contrary, dynamics similarities between inactive conformers are scattered among 3-4 (average value of 3.4) different directions, less shared among pairs of conformers. The corresponding KS statistic value between both distributions is 0.90 with a p-value= 3.9x10-11. That is, results to be slightly less effective than to separate active and inactive dynamics.

The identity of the two lowest common directions and for pairs of active conformers can be explored by analyzing their projections on the basis of the original normal modes. The average contribution of the 1st lowest normal mode to is 0.93±0.15 and the contribution of the 2nd lowest normal modes to is 0.92±0.15. Therefore, active conformers share two directions of motions mainly represented by the two lowest frequency normal modes. On the contrary, values of 0.59±0.38 and 0.29±0.32 are obtained for contributions of 1st and 2nd normal modes to and of inactive conformers. That is, while the dynamics of pairs of active conformers reveal common directions of motions that correspond to their two natural lowest frequency of vibration, the dynamics of pairs of inactive conformers do not present unique patterns that can be assigned to individual original normal modes.

As we have shown, pairs of active conformers share common directions of motion represented by directions and . In order to obtain a fingerprint that characterizes them among the complete set of active conformers, further analysis is required. SVD, as a data compression technique, highlights the main common features of the original directions within a few SVD representative modes . **Figure 7** depicts distributions of the overlap between SVD representative modes, , and original directions. As it can be seen, our complete set of active conformers share common directions represented by the corresponding SVD representative modes and . In contrast, the corresponding and in inactive conformers cannot be used as representative modes of the ensemble. Finally, directions are different among either pairs of active or inactive conformers. In summary, all active conformers of our dataset share common dynamics that can ultimately associated to their two lowest frequency modes. On the contrary, the dynamics of inactive conformers result heterogeneous and no unique dynamics patterns that include all of them can be found.

**Figure 8** shows that fingerprint modes and , shared by all active conformers, describe relative displacements between the two lobes (N and C lobes) of the kinase domain. Particularly, motions of β1-β2, β2-β3, β3-αC and β4-β5 loops on N lobe, and αG helix including the αG -αH loop on C lobe. Besides, a relative suppression of active loop deformations also characterizes modes and respect to motions of the corresponding modes of inactive conformations.

In order to evaluate effects of modes and on the active conformation of EGFR kinase domain we calculate the corresponding changes of radius of gyration (*R*g). **Figure 9** shows the distribution of values of *R*g due to displacements in the direction of these modes. As it can be seen, their effect on active conformers significantly differs from displacements of the corresponding modes of inactive conformers. Modes and of active conformers lead to more extended conformations involving a hinge-like separation of the N and C lobes (see also **Supplementary Info movies**).

The motion in the direction of and modes in active conformers is consistent with previously reported motions30,31 analyzed using molecular dynamics simulations, that describe the active-inactive conformational transition. Shan et al30 reported an opening of the two lobes to allow local unfolding at the hinge region, prior to the close of the lobes to re-stabilize in its inactive conformation. Our analysis suggests that these findings are actually a hallmark of active conformers dynamics when it comes to differentiate them from the inactive conformers.

**Yo creo que aca encajaria perfecto poder calcular el cambio de volumen de la cavidad debido al desplazamiento en la direccion del modo 1 y 2 en activas e inactivas….mostrando que en activas esos modos generan un cambio de volumen mucho mayor…te diria que pruebes eso. Si las 2 distribuciones dan separadas como suponemos, siendo el cambio de volumen de las activas mayor, pienso que amerita agregarlo ya que es una clara muestra del impacto de esas vibraciones en la funcionalidad…que se pierde en las inactivas. Pero para esto seria conveniente quizás publicar antes el método del calculo de volumen en Bioinformatics!!!! Sino la cita queda ambigua.**

## IV. CONCLUSIONS

We have performed an extensive comparative analysis of global dynamics features shared by existing conformers in EGFR kinase. We have identified clear directions of motions that can be used as fingerprints to differentiate active and inactive conformers. A novel procedure has been applied that can be used for sets of conformers in other proteins. The method allows the comparison of patterns of vibrations rescuing common representative directions of motions shared among the ensemble.

Two representative directions of motions have been identified and characterized among active conformations in EGFR kinase. These motions represent fingerprints of active conformers that can be added to previously reported structural features. Their conservation among the complete set of active conformers enlighten us on minima functional mechanisms within them. Inactive conformers have shown a general tendency to disrupt the collective motion of residues involved in these vibrations. It is therefore expected that functional aspects of EGFR kinase involve coordinated motions between residues that are mainly reflected in the two lowest frequency normal mode of active conformers. These modes allow active conformers to reach more extended conformations involving a hinge-like separation of the N and C lobes, a prior requirement to achieve a relatively stable inactive conformation. more stable inactive conformation.

ACA quizás Silvina complete con algún parrafo hablando de las posibilidades que da esta metodologia para que alguien que viene de la biología pueda saber como se movería su mutante y asi por ejemplo pensar que efecto tiene la mutacion en el desplazamiento de equilibrio conformacional o que inhibidores pueden ser mas adecuados

**Figure Captions**

**Figure 1.** (a) Active, and (b) inactive EGFR kinase conformers indicating the main structural elements that commonly distinguish them.

**Figure 2.** Human EGFR kinase domain. Resudes belonged to the the main pocket of the active site are pointed out (green) following the definition given by Hasenahuer et al.

**Figure 3.** Distributions of relative values of normalized accumulation of contributions of modes to B-factors *Bi* of residues belong to the main pocket of the active site of the EGFR kinase, i.e., *w*k/*w*1 (eq. 6).

**Figure 4.** Distribution of the fraction of residues involved in the motion of the two lowest frequency normal modes, calculated as values *Pk/N*, for active (a) and inactive (b) EGFR kinase respectively.

**Figure 5.** Histograms of backbone flexibility similarity *B*. The two histograms show the frequency distributions of *B* for active and a reference set of inactive conformers.

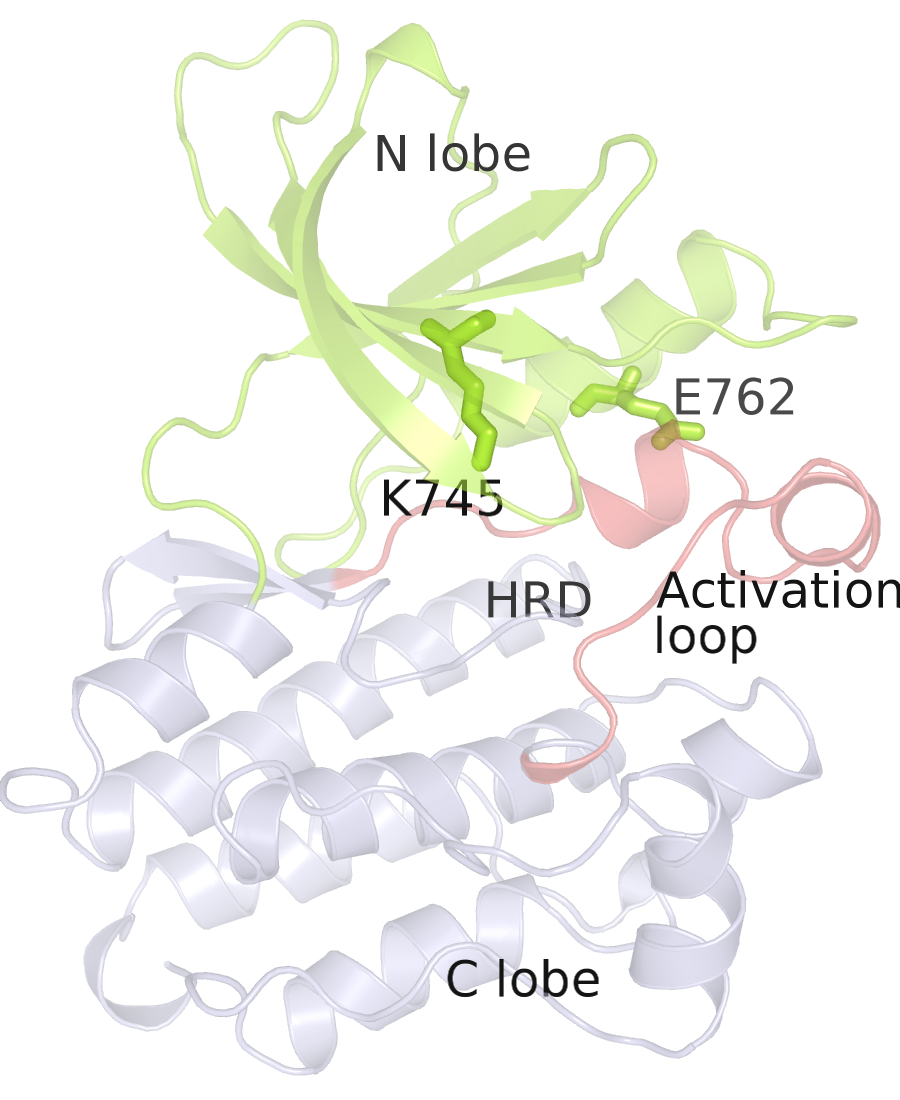
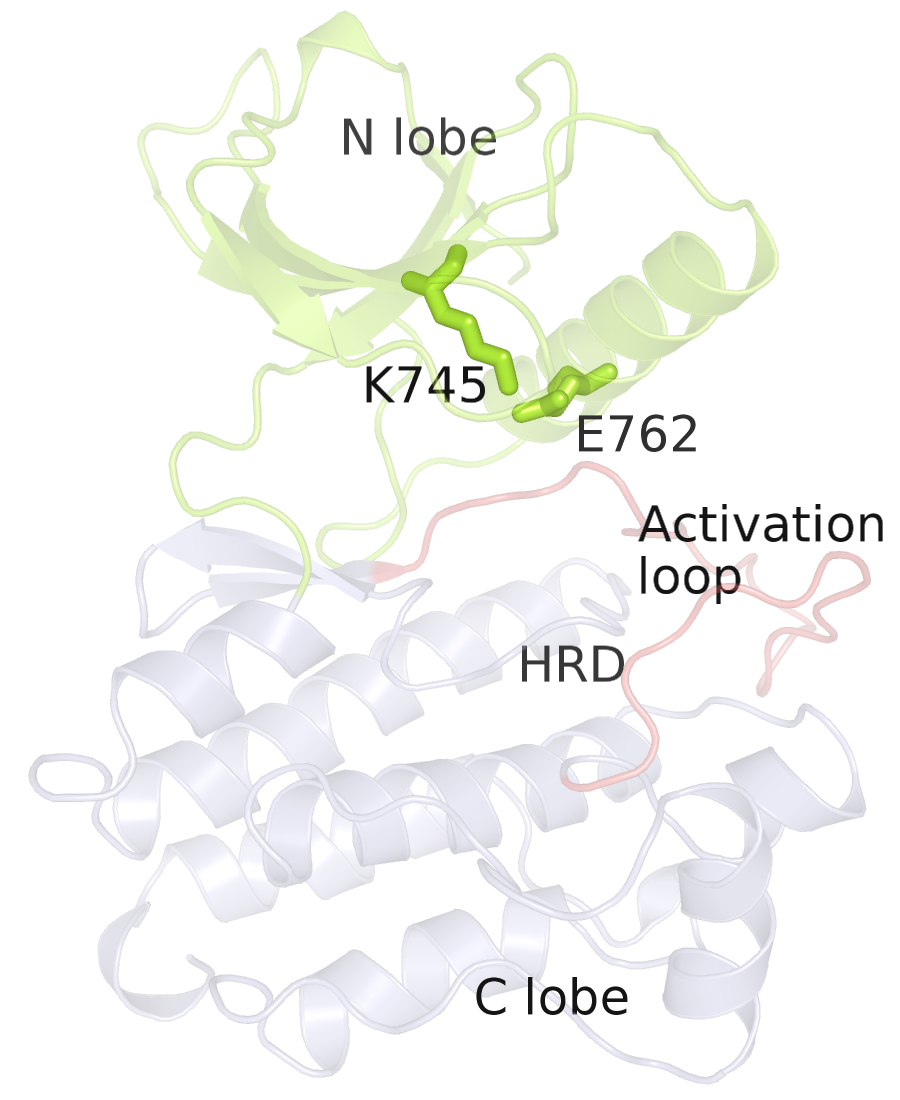
**Figure 6.** Histograms of values of **(a)** the similarity meassure between pairs of weighted vibrational spaces, and **(b)** the effective number of dimensions shared by pairs of structures calculated over all pairs of active and inactive conformers.

**Figure 7**. Histograms of the overlap between the first SVD representative modes, , and original directions for (a) ·; (b) ·; (c) · for both active (red) and inactive (blue) conformers.

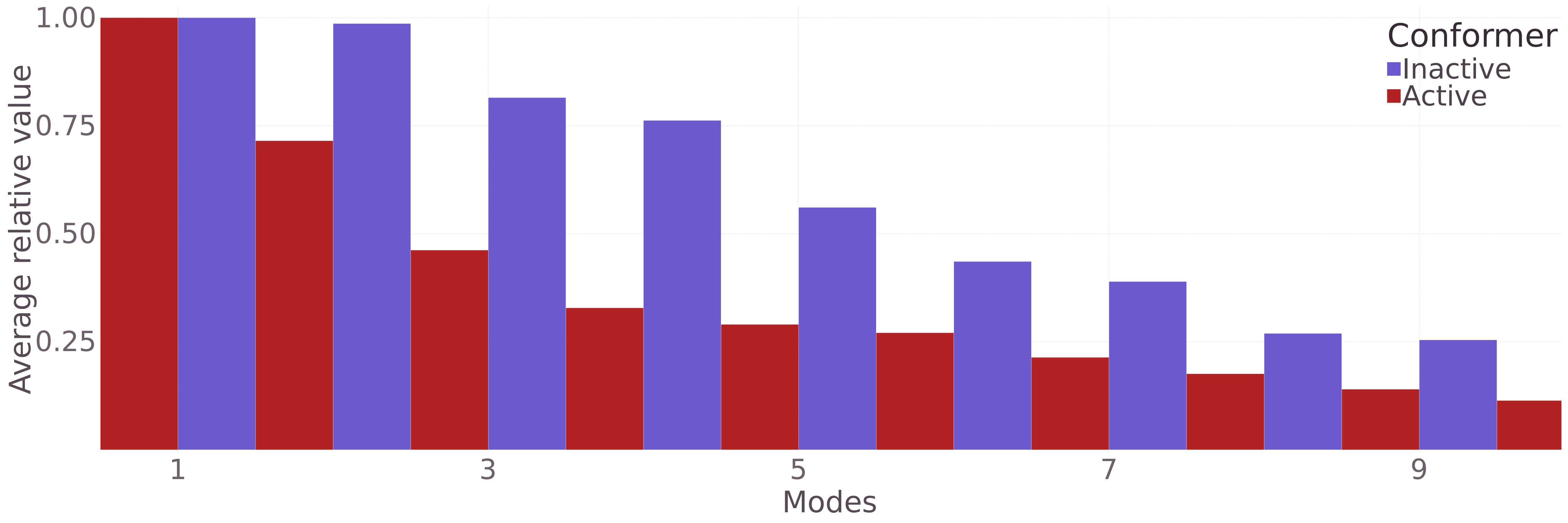
**Figure 8**. SVD representative modes (blue) and (green) of active conformers.

**Figure 9.** Histograms of changes of the radius of gyration, *R*g, of the EGFR kinase domain due to displacements in the direction of SVD representative modes (a) and (b) for both active (red) and inactive (blue) conformers.

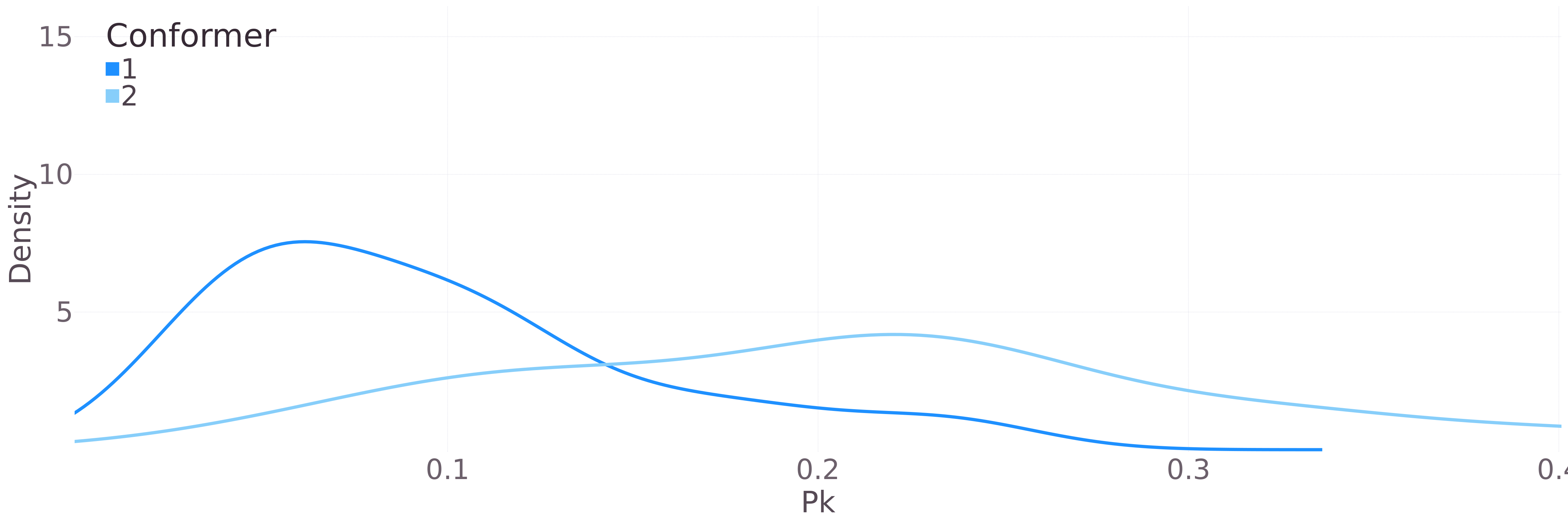
**Figure 1**

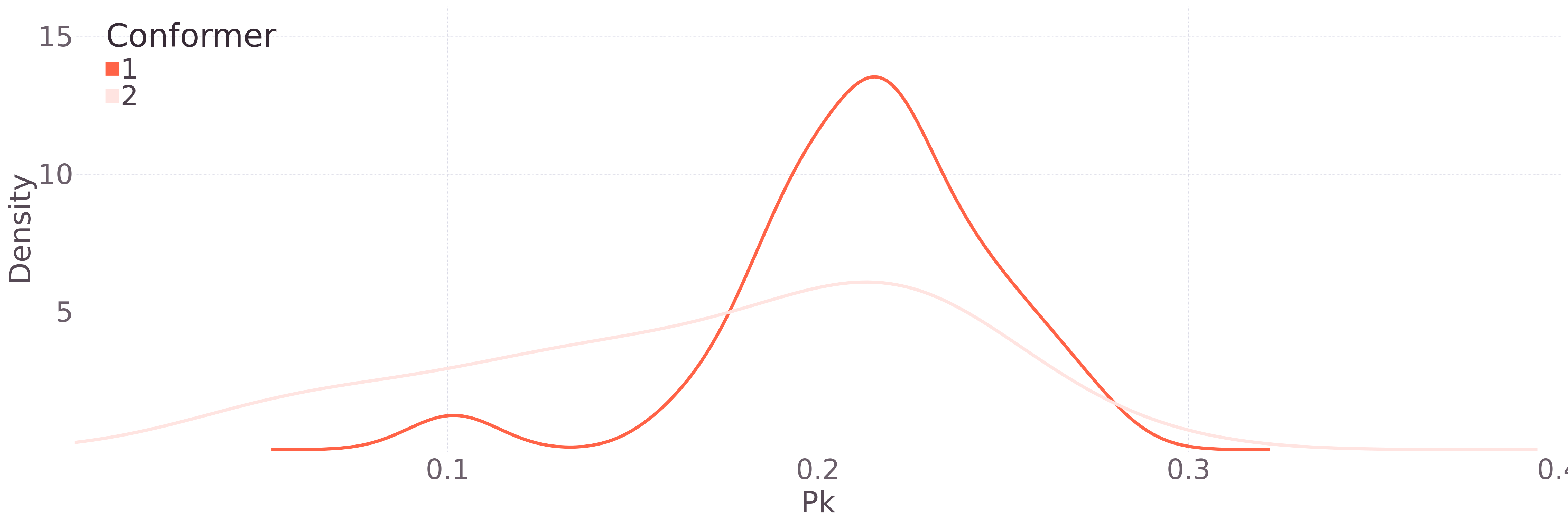
****

**Figure 3**

****

**Figure 4**

****

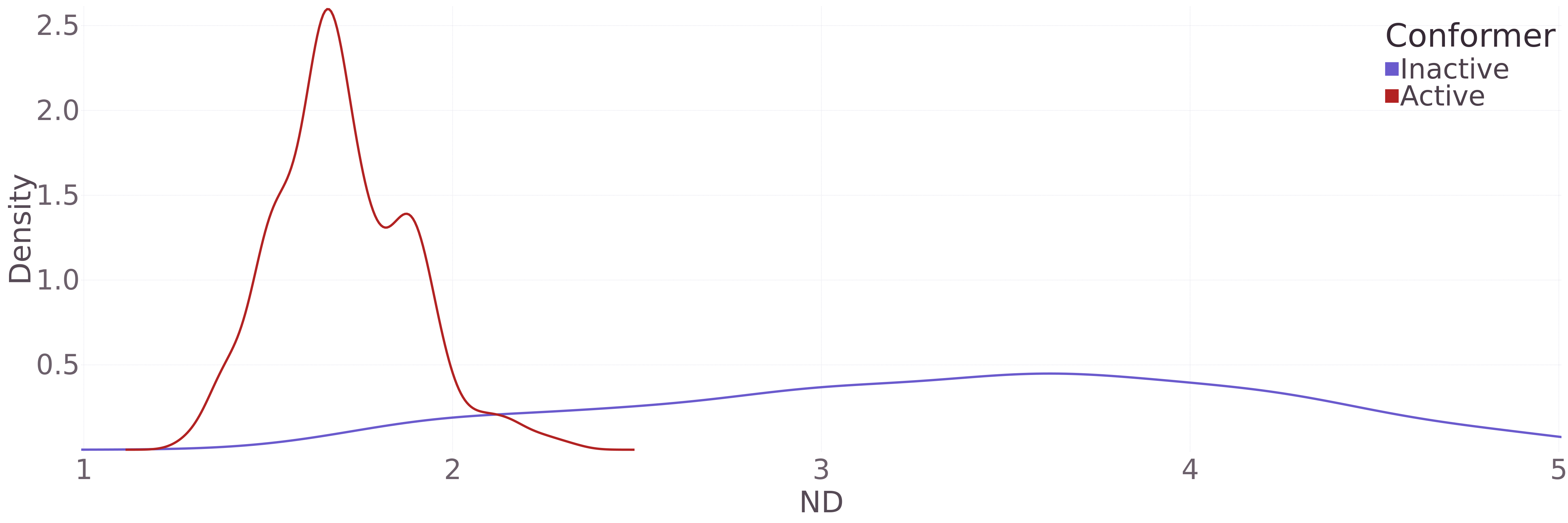
****

**Figure 5**

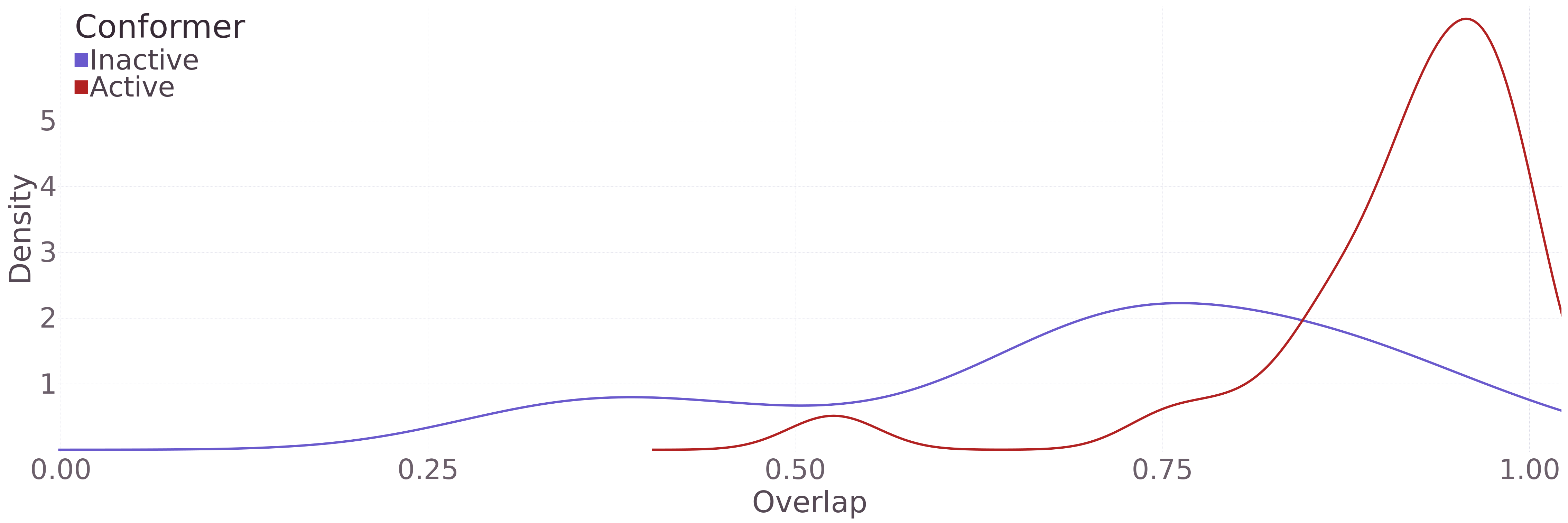
****

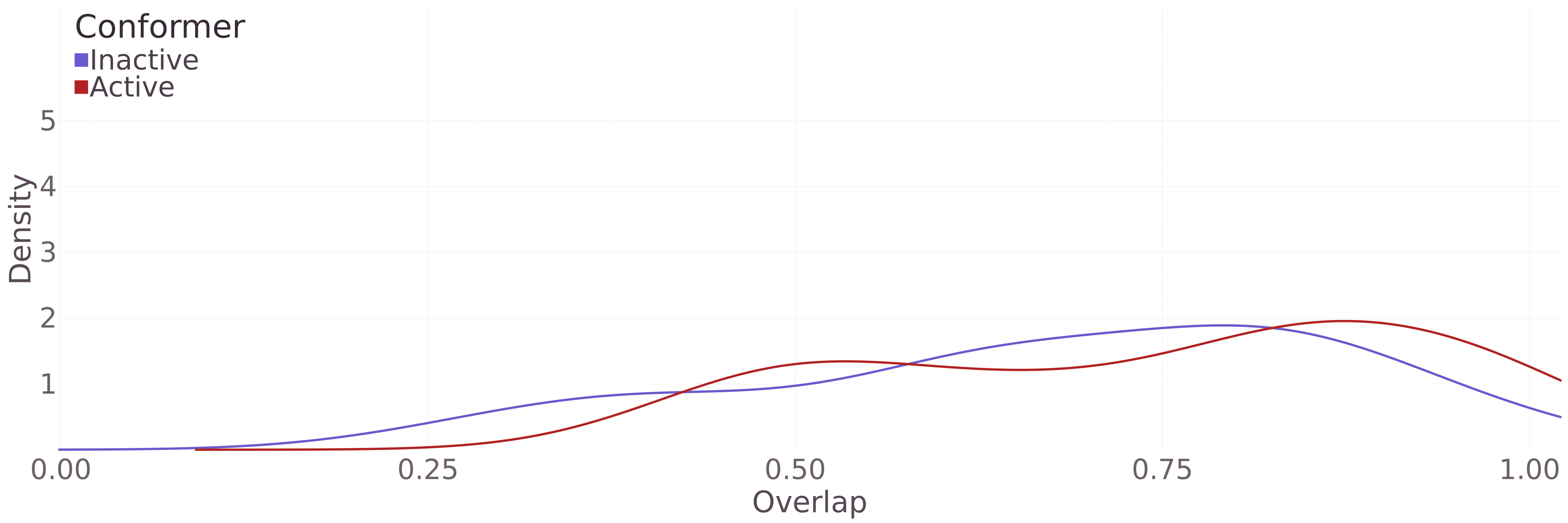
**Figure 6**

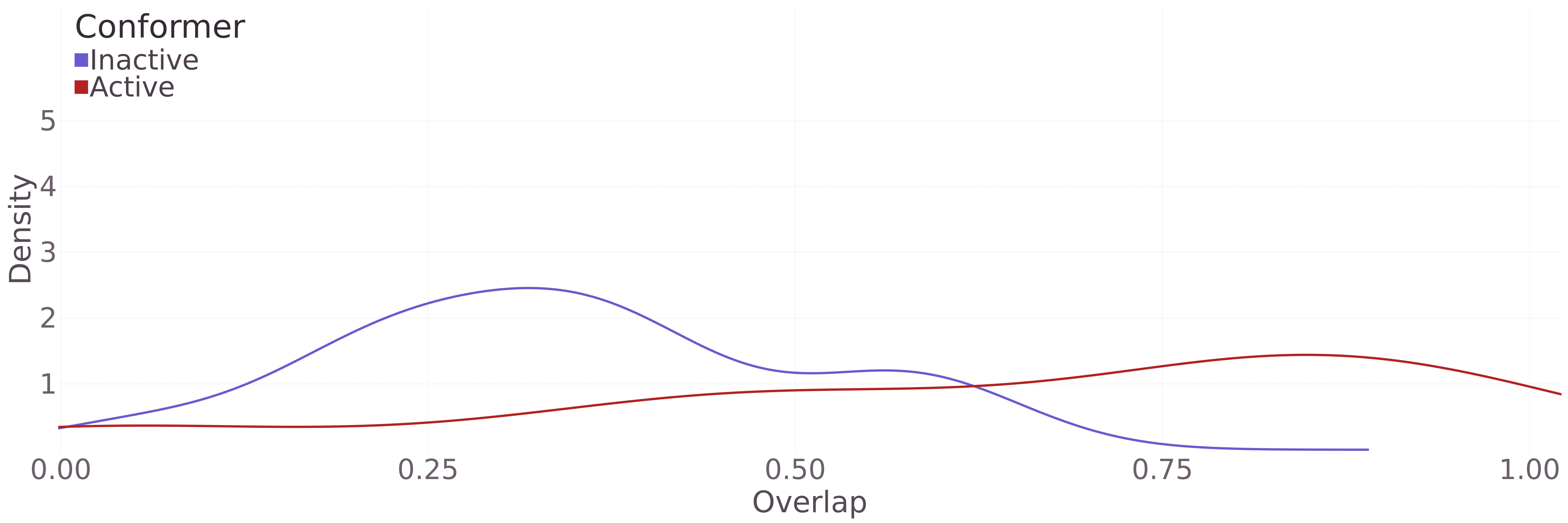
****

****

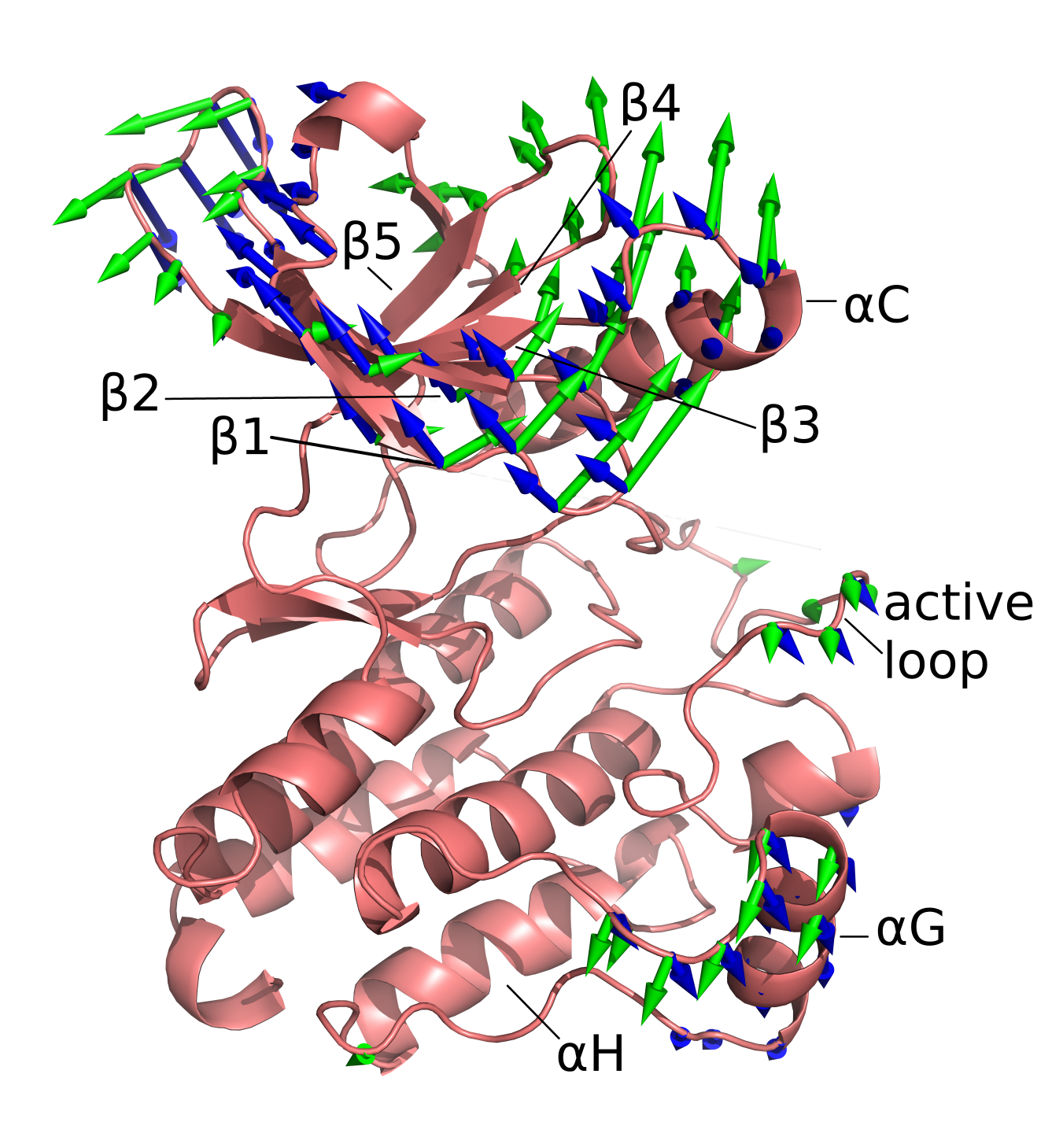
**Figure 7**

****

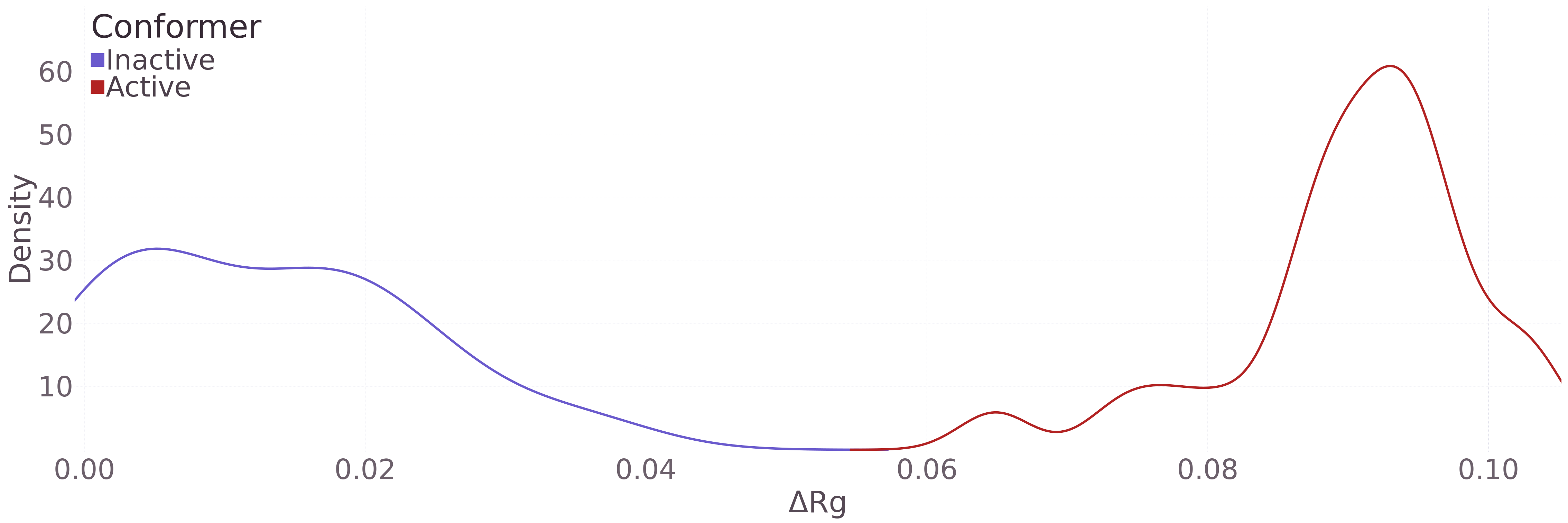
****

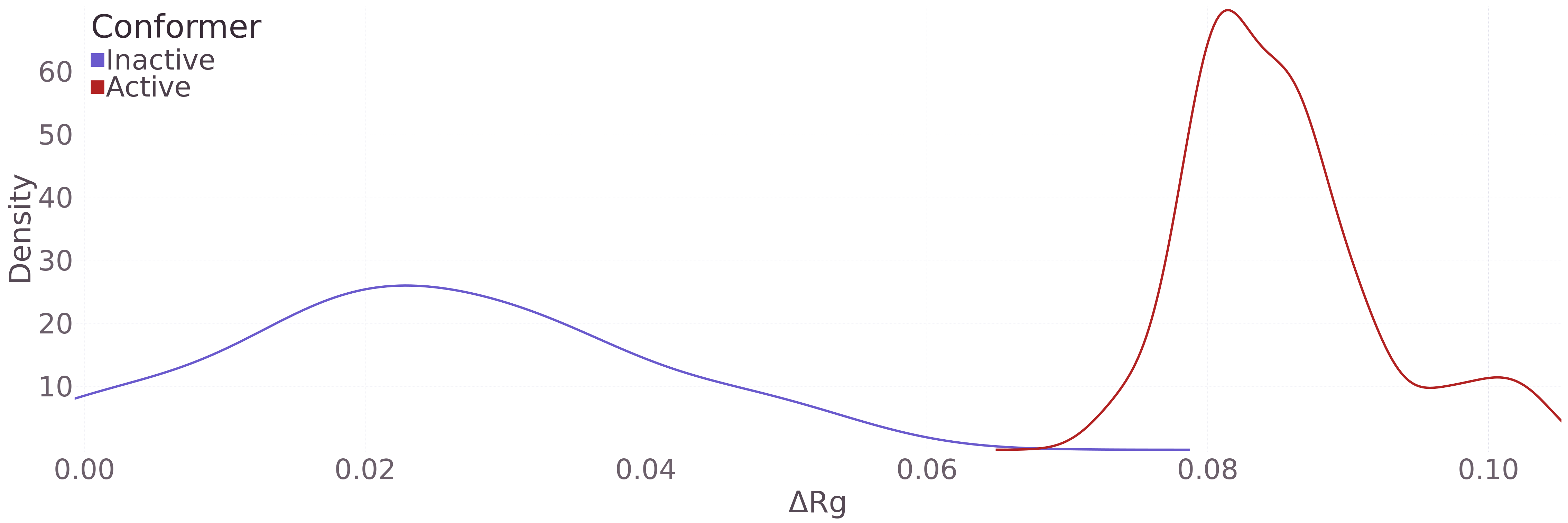
****

**Figure 8**



**Figure 9**

****

****

**References**

1 Y. Yarden and M. Sliwkowski, Nat. Rev. Mol. Cell. Biol. **2**, 127 (2001).

2 C. Arteaga and J. Engelman, Cancer Cell **25**, 282 (2014).

3 S. Forbes, D. Beare, and E. Al., Nucleic Acids Res. **43**, D805 (2015).

4 H. Cheng and E. Al, J. Med. Chem. **59**, 2005 (2016).

5 G. Lurje and H. Lenz, Oncology **77**, 400 (2009).

6 S. Müller and E. Al, Nat. Chem. Biol. **11**, 818 (2015).

7 Q. Wang, J. Zorn, and J. Kuriyan, Methods Enzymol. **548**, 23 (2014).

8 S. Kumar, B. Ma, C.J. Tsai, N. Sinha, and R. Nussinov, Protein Sci. **9**, 10 (2000).

9 S. Osorio, R. Alba, Z. Nikoloski, A. Kochevenko, A.R. Fernie, and J.J. Giovannoni, Plant Physiol. **159**, 1713 (2012).

10 L.C. James and D.S. Tawfik, **28**, 361 (2003).

11 X. Zhang, K. Pickin, R. Bose, N. Jura, P. Cole, and J. Kuriyan, Nature **450**, 741 (2007).

12 C. Yun, T. Boggon, Y. Li, M. Woo, H. Greulich, M. Meyerson, and M. Eck, Cancer Cell **11**, 217 (2007).

13 M. Kalia and K. Madhy, Metabolism. **64**, S16 (2015).

14 K. Fergurson, Biochem. Soc. Trans. **32**, 742 (2004).

15 A. Kornev and S. Taylor, Biochim. Biophys. Acta **1804**, 440 (2010).

16 N. Jura and E. Al., Mol. Cell **42**, 9 (2011).

17 A. Dixit and G. Verkhivker, Comput. Math. Methods Med. 653487 (2014).

18 D. Knighton and E. Al., Science (80-. ). **253**, 407 (1991).

19 W. Hemmer and E. Al., J. Biol. Chem. **272**, 16946 (1997).

20 S. Taylor and A. Kornev, Trends Biochem. Sci. **36**, 65 (2011).

21 K. James and G. Verkhivker, PLoS One **9**, e113488 (2014).

22 J. Hu and E. Al., Mol. Cell. Biol. **35**, 264 (2015).

23 K. Gajiwala and E. Al, Structure **21**, 209 (2013).

24 D. Fabbro, Mol. Pharmacol. **87**, 766 (2015).

25 M.A. Hasenahuer, G.P. Barletta, S. Fernandez-Alberti, G. Parisi, and M.S. Fornasari, PLoS One **12**, e0189147 (2017).

26 S. Kumar and R. Nussinov, J. Mol. Biol. **293**, 1241 (1999).

27 A. Tramontano and T. Anna, FEBS Lett. **580**, 2928 (2006).

28 H. Du and E. Al., **71**, 77 (2015).

29 M. Hasenahuer and E. Al, Ann. Hum. Genet. **79**, 385 (2015).

30 Y. Shan, A. Arkhipov, E.T. Kim, A.C. Pan, and D.E. Shaw, Proc. Natl. Acad. Sci. U. S. A. **110**, 7270 (2013).

31 Y. Shan, M.P. Eastwood, X. Zhang, E.T. Kim, A. Arkhipov, R.O. Dror, J. Jumper, J. Kuriyan, and D.E. Shaw, Cell **149**, 860 (2012).

32 M.M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).

33 I. Bahar, B. Erman, R.L. Jernigan, A.R. Atilgan, and C.D. G, J. Mol. Biol. **285**, 1023 (1999).

34 I. Bahar and R.L. Jernigan, Biochemistry **38**, 3478 (1999).

35 K. Hinsen and R.G. Kneller, J. Chem. Phys **24**, 10766 (1999).

36 A. Emperador, O. Carrillo, M. Rueda, and M. Orozco, Biophys. J. **95**, 2127 (2008).

37 M. Levitt, C. Sander, and P.S. Stern, J. Mol. Biol. **181**, 423 (1985).

38 O. and S.Y. Marques, Proteins **23**, 557 (1995).

39 C.Y. Xu, D.. Tobi, and B. I, J. Mol. Biol. **333**, 153 (2003).

40 L. Yang, G. Song, and R.L. Jernigan, Biophys. J. **93**, 920 (2007).

41 O. Keskin, R.L. Jernigan, and I. Bahar, **78**, 2093 (2000).

42 A. Leo-Macias, P. Lopez-Romero, D. Lupyan, D. Zerbino, and A.R. Ortiz, Biophys. J. **88**, 1291 (2005).

43 W. Zheng, B.R. Brooks, and D. Thirumalai, Proc. Natl. Acad. Sci. U. S. A. **103**, 7664 (2006).

44 S. Maguid, S. Fernandez-Alberti, L. Ferrelli, and J. Echave, Biophys. J. **89**, 3 (2005).

45 S. Maguid, S. Fernandez-Alberti, and J. Echave, Gene **422**, 7 (2008).

46 T. Saldaño, A. Monzon, G. Parisi, and S. Fernandez-Alberti, PLOS Comput. Biol. **12**, e1004775 (2016).

47 A.M. Monzon, E. Juritz, M.S. Fornasari, and G. Parisi, Bioinformatics **29**, 2512 (2013).

48 A. Sali and T. Blundell, J. Mol. Biol. **234**, 779 (1993).

49 K. Hinsen, Proteins Struct. Funct. Genet. **33**, 417 (1998).

50 a R. Atilgan, S.R. Durell, R.L. Jernigan, M.C. Demirel, O. Keskin, and I. Bahar, Biophys. J. **80**, 505 (2001).

51 J.I. Jeong, Y. Jang, and M.K. Kim, J. Mol. Graph. Model. **24**, 296 (2006).

52 V. Sobolev, a Sorokine, J. Prilusky, E.E. Abola, and M. Edelman, Bioinformatics **15**, 327 (1999).

53 I. Bahar, a R. Atilgan, and B. Erman, Fold. Des. **2**, 173 (1997).

54 F. Tama, F.X. Gadea, O. Marques, and Y.H. Sanejouand, Proteins Struct. Funct. Bioinforma. **41**, 1 (2000).

55 W.J. Krzanowski, J. Am. Stat. Assoc. **74**, 703 (1979).

56 M. Grosso, A.. Kalstein, G. Parisi, A. Roitberg, S. Fernandez-Alberti, M. Grosso, A. Kalstein, G. Parisi, and A.E. Roitberg, J. Chem. Phys. **142**, 245101 (2015).

57 M. Kirkpatrick, Genetica **136**, 271 (2009).

58 M.E. Wall, A. Rechtsteinner, and L.M. Rocha, *Singular Value Decomposition and Principal Component Analysis. In A Practical Approach to Microarray Data Analysis.* (Kluwer, Norwell, MA., 2003).