Chemometrical Classification of Ephrin Ligands and Eph Kinases Using GRID/CPCA Approach

Eugene Myshkin and Bingcheng Wang*

Rammelkamp Center for Research, MetroHealth Campus, and Department of Pharmacology and Ireland Cancer Center, Case Western Reserve University School of Medicine, 2500 MetroHealth Drive, Cleveland, Ohio 44109

Received December 19, 2002

Eph receptor tyrosine kinases are divided on two subfamilies based on their affinity for ephrin ligands and play a crucial role in the intercellular processes such as angiogenesis, neurogenesis, and carcinogenesis. As such, Eph kinases represent potential targets for drug design, which requires the knowledge of structural features responsible for their specific interactions. To overcome the existing gap between available sequence and structure information we have built 3D models of eight ephrins and 13 Eph kinase ligand-binding domains using homology modeling techniques. The interaction energies for several molecular probes with binding sites of these models were calculated using GRID and subjected to chemometrical classification based on consensus principal component analysis (CPCA). Despite inherent limitations of the homology models, CPCA was able to successfully distinguish between ephrins and Eph kinases, between Eph kinase subfamilies, and between ephrin subfamilies. As a result we have identified several amino acids that may account for selectivity in ephrin−Eph kinase interactions. In general, although the difference in charge between ephrin and Eph kinase binding domains creates an attractive long-range electrostatic force, the hydrophobic and steric interactions are highly important for the short-range interactions between two proteins. The chemometrical analysis also provides the pharmacophore model, which could be used for virtual screening and de novo ligand design.

INTRODUCTION

The development of multicellular organism requires intricate intercellular communications, mediated among other proteins by receptor tyrosine kinases (RTK). The largest family (25%) of RTKs is constituted by the Eph kinases,[1] consisting of 15 members and divided into two subfamilies based on sequence homology and their affinity for the ligands. EphA kinases bind to glycosylphosphatidylinositol (GPI) anchored ligands (ephrinA), while EphB kinases target ligands with transmembrane domains (ephrinB)[2] (http://cbweb.med.harvard.edu/eph-nomenclature). The prerequisite for the membrane attachment of ligands distinguishes Eph kinases from other known RTKs, like EGF receptor. Eph kinases are composed of several domains: ligand-binding globular domain, cysteine-rich region, and two fibronectin type III repeats constitute an extracellular part. Juxtamembrane region, protein kinase domain, a sterile α-motif (SAM), and PDZ domain comprise the cytoplasmic part of the Eph kinases and initiate forward signaling.[3] The biochemical evidence suggests that ephrin ligands are also capable of signal transduction.[4,5] This bidirectional signaling,[6] which is crucial for cell−cell adhesion, spreading, and migration, is another unique characteristic of this RTK family. It was shown that Eph kinases play a major role in diverse developmental processes,[7] including angiogenesis[8] and axon pathfinding.[9] They are also involved in tumorigenesis.[10] We have found that EphA2 kinase activation in prostate cancer cells inhibits cell migration and proliferation through novel signal pathways.[11,12] As such, Eph kinases represent potential targets for drug discovery.

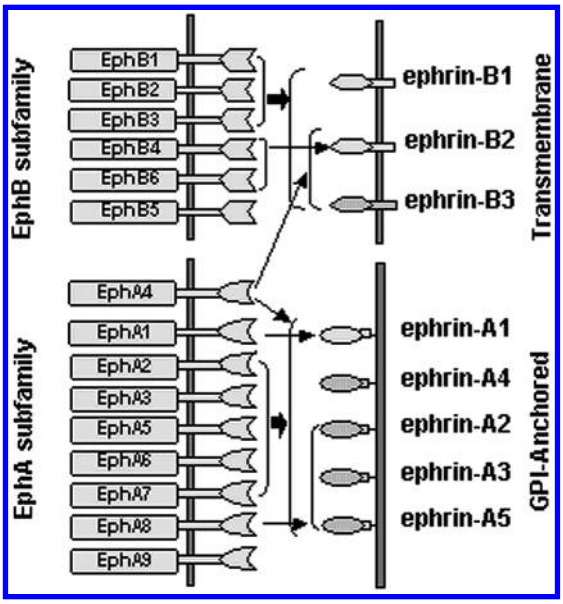

Figure 1. Ephrin and Eph kinase interaction network scheme. Note the selective interactions between Eph kinases and ephrins of the same class. EphB4 interacts only with epnrinB2, EphA1 only with ephrinA1, and EphA4 with the ligands of the B class.

The experimentally derived ephrin−Eph kinase interaction network is represented in Figure 1. It was determined that the Eph receptors promiscuously bind to ephrin ligands

* Corresponding author phone: (216)778-0256; e-mail: bxw14@po.cwru.edu. Corresponding author address: 2500 MetroHealth Drive, Cleveland, OH 44109.

10.1021/ci0256586 CCC: $25.00
Published on Web 03/21/2003

within the same class. The exceptions are EphA4 kinase, which, besides ephrinA, binds also to ephrinB2 and ephrinB3, EphA1 that reacts only with ephrinA1 and EphB4 kinase, which binds only to ephrinB2. The ligand specificity of EphB5 is not yet clear. The understanding of ephrin−Eph kinase interactions requires the knowledge of structural features that distinguish between two subfamilies and are responsible for their specific interactions. This knowledge can also be utilized in the design of small molecules that selectively target the receptors of interest. The ligand selectivity is crucial for the minimization of potential side effects and should be addressed on the earliest stages of the drug design process. In this paper we are using GRID/CPCA[13] approach to characterize the binding sites of ephrin ligands and Eph receptors, which will allow us to discern the structural features responsible for their specific interactions and lead to the design of the selective small molecules.

GRID/CPCA approach has already been proved valid for the analysis of the binding sites of thrombin proteases,[13] cytochromes,[14] metalloproteases,[15] and structural classification of the protein tyrosine kinase domains.[16] Here we focus on the binding sites of ephrins and Eph kinases. To date only crystal structures of ehprinB2,[17] EphB2,[18] and their complex[19] are available. In the absence of crystal structure, comparative modeling techniques,[20] as implemented in program MODELLER,[21,22] can provide reliable 3D models for the proteins of interest. The molecular interaction fields around the binding sites of the homology models are calculated with program GRID[23,24] and then used as 3D-descriptors for the chemometrical analysis, which can classify the Eph kinases and their ephrin ligands based on their interactions with GRID molecular probes. The energy contributions from molecular probes can be readily visualized as energy isocontours simplifying the analysis of the binding sites. The estimation of these energies can be overlooked just using the simple inspection of homology models. The chemometrical tools used for multivariate classification are based on principal component analysis (PCA)[25] and consensus PCA (CPCA)[26] and implemented in program GOLPE.[27,28] The advantage in using CPCA is that it also evaluates the contribution of different molecular probes to the selectivity. Overall, this study is meant to investigate protein−protein interactions between ephrins and Eph kinases and to overcome the existing gap between available sequence and structure information, using state of the art computational tools.

COMPUTATIONAL METHODS

Protein Structures. The crystal structure of the ephrin−Eph kinase complex was used as a template for the homology modeling.[19] The sequences of ephrins and Eph kinases were retrieved from the GenBank. The program MODELLER,[29] which gives better results in modeling loops, multiple conformation states, and problematic sequence alignments, was used for homology modeling. The ALIGN2D module of MODELLER, that utilizes global dynamic programming algorithm with higher gap penalty for insertions/deletions in structurally conserved regions, was used for sequence alignment. The alignments were adjusted manually where necessary. Several homology models for each structure were generated, using default values in the program. MODELLER calculates spatial restraints such as Cα−Cα distances, dihedral angles from the template structure, and attempts to fit the target sequence by satisfying this restraints using molecular dynamics. The structures with the best score represented by the lowest value of MODELLER objective function were selected and subjected to energy minimization. Then the structures were visually inspected, and the stereochemistry of the models was evaluated using PROCHECK program.[30,31] The root mean standard deviation (RMSD) of ephrin and Eph kinase 3D-models from ephrinB2 and EphB2 structures, respectively, were calculated with SwissPDB-Viewer[32] by superimposing Cα carbons of the backbone. The binding sites were analyzed for hydrogen bonding and hydrophobic interactions by HBPLUS[33] and the DIMPLOT module of the LIGPLOT software.[34] ClustalX was also used for sequence alignment.[35]

Table 1. List of GRID Probes and Their Properties Used in This Study

GRID probe	chemical group
DRY	hydrophobic probe (entropic term)
C3	Sp3 methyl probe (steric term)
N1	neutral flat NH (e.g. amide − polar)
O	Sp2 carbonyl oxygen (polar)

GRID Calculations. The molecular interaction fields around the binding sites were calculated with GRID20. The proteins were considered rigid (directive MOVE=0). The grid box dimensions were scaled to include the binding sites of ephrinB2 and EphB2 in the dimer structure yielding a 37 Å × 28 Å × 44 Å box. The grid spacing was 1 Å, and the default values for the other directives were used. Hydrogens were added to the structures with the GRIN module. The molecular probes (Table 1) accounting for hydrophobic, steric, and polar interactions were selected for the field calculations.

Chemometrical Analysis. GRID fields were imported in GOLPE and organized in the table with rows corresponding to the objects (proteins) and columns corresponding to the variables. The variables are the values of the interaction field for a given probe and grid point and are arranged in a linear vector concatenated with analogous vectors obtained from other probes. The details are described elsewhere.[11] Before multivariate analyses the data were subjected to the pretreatment as follows. All the data with values less than 0.01 kcal/mol and standard deviation less than 0.02 kcal/mol were set to zero. To facilitate the interpretation of the results the positive values of energy from unfavorable interactions were eliminated by setting maximum positive cutoff to zero. To reduce the number of variables and facilitate the computation the region cutout tool was used. Only the data with 4 Å of the binding sites were left for analyses. The block unscaled weights (BUW) procedure was applied to scale the importance of the different probes. The resulting data matrix is subjected to the PCA, which reduces this multivariate space into few principal components (PC) describing most of the variance in the data. Essentially, PCA decomposes data matrix on the product of score and loading matrices. The scores represent the relationship of PC to the objects and the loadings represent the relationship to the variables. Here we used CPCA, which belongs to the hierarchical PCA and is useful when the variables are clustered in blocks corresponding to individual probes. In this case it also calculates

Table 2. Summary of Structural Statistics for Ephrins and Eph Kinases

Eph kinases	charge[a]	homology[b]	RMSD[c]	ephrin ligands	charge	homology	RMSD
A1	0	43	0.57	a1	−2	38	0.61
A2	−7	47	0.47	a2	0	35	0.66
A3	−5	59	0.41	a3	5	39	0.59
A4	−6	63	0.37	a4	−4	35	0.63
A5	−8	56	0.34	a5	1	36	0.74
A6	−4	59	0.31	b1	4	63	0.28
A7	−4	61	0.37	b2	1	100	0
A8	−3	54	0.43	b3	1	56	0.41
B1	−2	77	0.24				
B2	−1	100	0				
B3	−4	72	0.35				
B4	−1	46	0.36				
B5	−6	63	0.33				

[a] Charge was calculated using GRID. [b] Homology is calculated as pairwise sequence identity (%). [c] RMSD − root mean standard deviation of model structure from crystal structure.

block scores and block loadings, providing unique information about the importance of a particular block (probe) for PC.

RESULTS

Homology Models. The first step in homology modeling is the template selection. For this family of RTK only the structures of ephrin B2, EphB2, and their complex are available with EphA2 structure under way (Nikolov and Wang, unpublished). Since the binding of two proteins involves specific conformational changes and because our analysis mainly focuses on the binding interfaces of ephrins and Eph kinases, the ephrinB2−EphB2 complex structure was selected as a template for our comparative modeling studies. We have used all the ephrins and most of the Eph kinases as the targets sequences, except EphB6 kinase, which has a short insertion in the E−F loop and polyserine insertion in the J−K loop, that make this protein hard to model.

The second step, which is the most crucial for the building of the correct homology model, is the target-template alignment. The accuracy of the modeled structure depends on the extent of homology between the proteins. For the protein with homology higher than 40% the alignment is almost always correct.[36] To this end we have calculated pairwise sequence identity (SI) for the receptor-binding domain of ephrins and ligand-binding domain of Eph receptors using the PredictProtein server (http://www.embl-heidelberg.de/predictprotein/submit_adv.html) (Table 2). These calculations suggest that medium to high accuracy models can be obtained. It is interesting that the least homologous to EphB2 is EphA1 with 43% SI. The highest homology for EphB2 have EphB1 and EphB3 with 77% and 72% SI, respectively. Surprisingly, the third kinase with high homology is EphA4 with 63% SI. This fact alone may account for its unique reactivity toward not only ephrinA but also ephrinB ligands. It is also quite unique that EphB4, which interacts only with ephrinB2 and not ephrinB1 and ephrinB3, is the least homologous to EphB2 with 46% SI compared to other EphB kinases. The least homologous to ephrinB2 is ephrinA1 which has 35% SI. It should also be noted that the ligand-binding domains of Eph kinases are highly negative and ephrins are mostly positive in charge (Table 2).

As a result of relatively high homology all obtained models have a conserved fold and low RMSD values of Cα atoms (Table 2) characteristic for the high accuracy models. The quality of the models was evaluated by stereochemical analysis using PROCHECK. From the analysis of the Ramachandran plots it was observed that most of the amino acids are clustered in favored or allowed regions. Only one or two amino acids on average for Eph kinases are located in the disallowed regions. Ephrins have about four amino acids in the disallowed regions. As expected, this highly depends on homology of the targets to the template sequence. Higher number of the disallowed amino acids is the result of the conserved deletion in the C-terminal region of ephrinA class and lower SI. For all the models the main-chain parameters, like bad contacts or overall G-factor, and side-chain stereochemical parameters are better or inside the regions statistically derived from analysis of crystal structure databases. These results for Eph kinases and ephrins are in agreement with those of crystal structures of EphB2 and ephrinB2, respectively, subjected to the same analysis.

Chemometrical Classification. The obtained 3D-models were subjected to chemometrical classification. The data collection was performed with program GRID, which was used for the calculation of molecular interaction fields for three sets of molecules: set of Eph kinases, set of ephrins, and set of ephrins and Eph kinases. The grid was focused on the binding sites. Preliminary CPCA suggested that only four probes (Table 1) are relevant for explaining most of the variance in PC and do not provide similar information. A hydrophobic component plays a major role in protein interactions,[37] but due to the smaller values of energy its contribution can be diminished when considered together with electrostatic interactions. To normalize their importance the data were pretreated with BUW. The energy distributions for four GRID probes before and after pretreatment are shown on Figure 2. It can be seen that after pretreatment the DRY probe exhibits higher interaction energies for Eph kinases than other probes.

The plot of object scores against the PCs represents a map of these objects. Each dot on this plot corresponds to a particular protein structure. Proteins with similar properties will be clustered in the same regions of this map. Usually, first two PCs describe the most variance in the data and can be easily visualized, so they will be used in our analysis. The differential PC score plot using all for GRID probes (Table 1) for the set of ephrins and Eph kinases is represented in Figure 3A. It can be observed that the first PC effectively discriminates between cluster of ephrins and a cluster of Eph kinases, located at the opposite sites of this PC, which is in agreement with complementary nature of interactions between these proteins. The second and third PCs (not shown) distinguish between Eph kinases and are essentially the same as the first two PCs for the Eph kinase set (Figure 3B).

On Figure 3B the first PC separates Eph kinases from three outlier objects (EphB4, EphA1, and EphA2). These outliers have biological significance and thus do not represent errors of the model. As mentioned earlier, EphB4 interacts only with ephrinB2, which is unique for EphB subclass, and EphA1 interacts only with ephrinA1, which is unique for EphA subclass. The score plot for Eph kinases calculated without outliers is depicted in Figure 3C. On this plot the

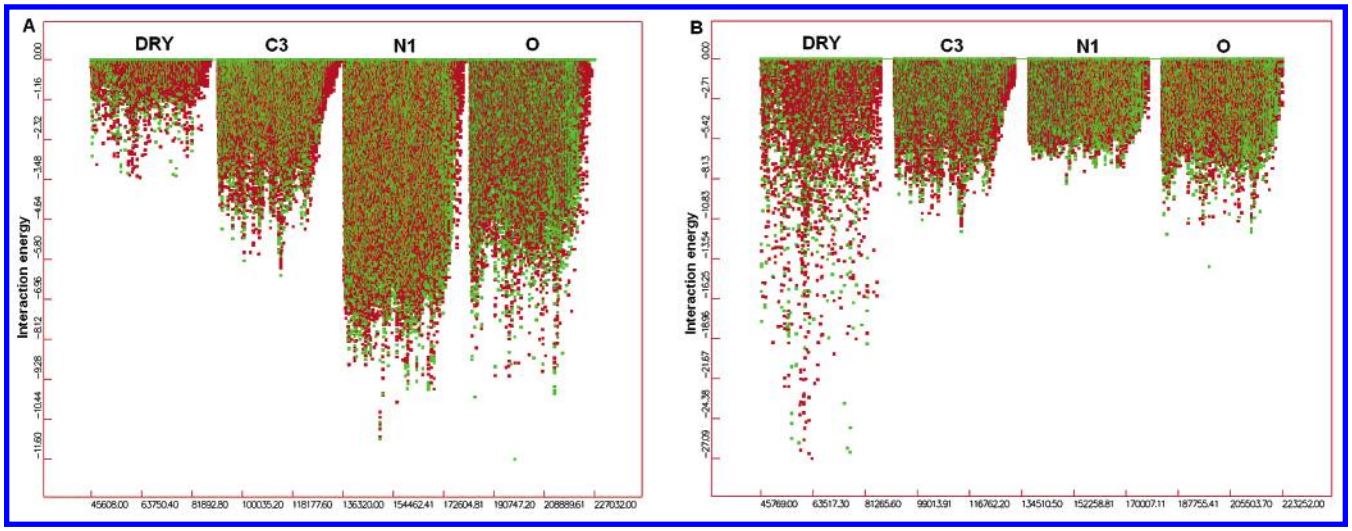

Figure 2. Distribution of interaction energies for four GRID probes (DRY, C3, N1, O) before (left) and after (right) BUW pretreatment. Red dots represent EphA kinases, green dots − EphB kinases. Note the importance of the hydrophobic probe for the protein−protein interactions after pretreatment.

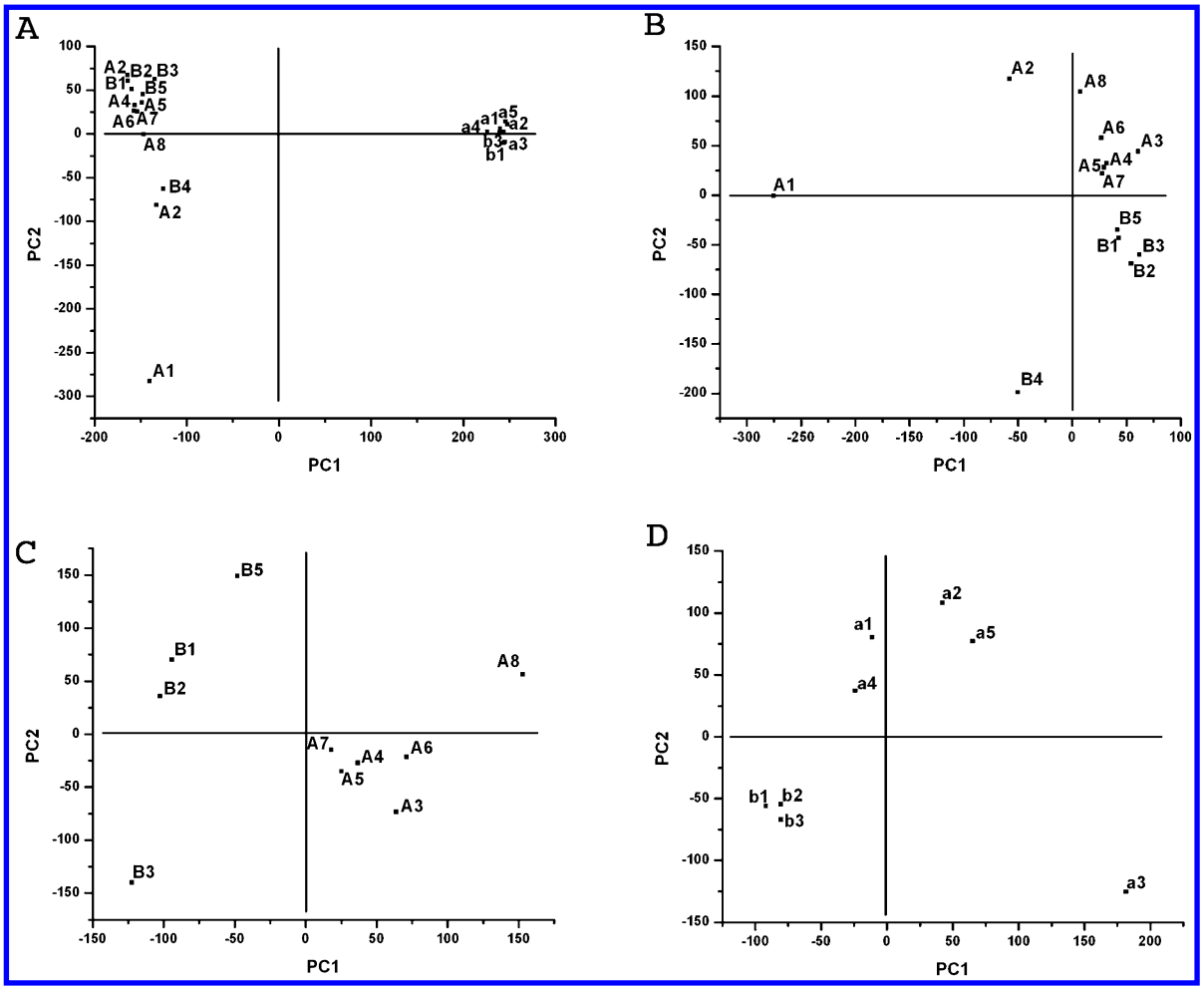

Figure 3. PCA differential scoreplots using all four GRID probes for the set of ephrins and Eph kinases (A), the set of Eph kinases (B), Eph kinases without outliers (C), and the set of ephrins (D). The plots illustrate the difference between Eph and ephrin subfamilies. Eph kinases are in uppercase, ephrins − in lowercase.

first PC separates EphA and EphB subclasses. EphB kinases are spread along the second PC, which also distinguishes EphA8 with different interaction pattern (Figure 1), from the cluster of other EphA kinases. Surprisingly, CPCA was unable to reveal unique properties of EphA4 kinase, which would be expected to locate on the border between two groups. The score plot for the set of ephrins in Figure 3D separates the less spread cluster of ephrinBs from ephrinAs. In general, PCA based on calculated molecular interaction fields for 3D models effectively separates ephrins and Eph kinases on two classes which have relevant biological significance.

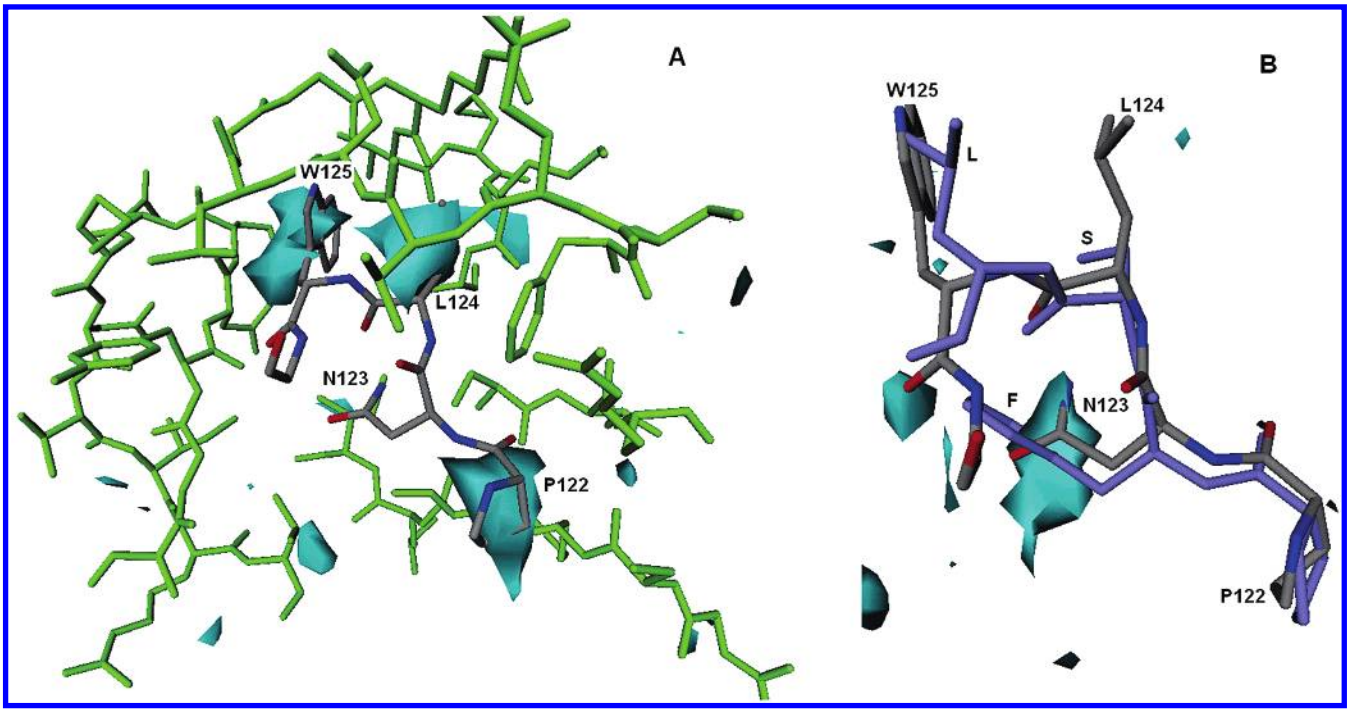

Figure 4. (A) CPCA pseudofield plot for the C3 probe differentiating between Eph kinases and ephrins. Negative contours (in cyan) show favorable steric interactions within Eph binding site. Only four amino acids of G−H ephrinB2 loop are shown within the EphB2 binding site (in green). (B) CPCA pseudofield plot for the O probe differentiating between EphA and EphB kinases. Negative contours (in cyan) show favorable polar interaction for EphB kinase. Only four amino acids of G−H ephrinB2 loop (atom properties color) and homologous amino acids of ephrinA2 (in purple) are shown. The numeration is that of ephrinB2.

Analysis of Ephrin−Eph Interactions. To investigate the structural basis of ephrin−Eph kinase interactions we used active plot option in GOLPE, where the difference between two points on the score plot can be translated back into the original multivariate space through PC loadings. As a result a differential pseudofield plot is obtained. This plot is represented as energy isocontours, demonstrating what regions of the binding site are preferable or selective for one object with respect to the other and what interaction is responsible for this selectivity. To facilitate the analysis we will focus mostly on the interaction of G−H loop in ephrins with the cleft formed by D−E and J−K loops of Eph kinases.

The C3 probe differential plot for the cluster of Eph kinases and the cluster of ephrins is shown in Figure 4A. Only negative energy contours (in cyan), representing the regions favorable for interaction of C3 group with Eph kinases, are shown here. The numeration of residues is based on ephrinB2 and EphB2 since their structures are readily available. The C3 probe describes the steric and hydrophobic interactions within the binding site. As can be seen from this contour plot, these favorable steric regions in Eph kinases are precisely targeted by three hydrophobic amino acids, Trp125, Leu124, and Pro122, which belong to G−H ephrin loop. Among them Pro122 is a highly conserved residue in all ephrins, except ephrinA3 (Figure 5). This fact may also account for a separate position of ephrinA3 on the ephrin score plot (Figure 3D). The aromatic Trp125 in ephrins B2 and B3, which also forms a hydrogen bond with glycine in EphB2 kinase, is replaced by aliphatic leucine conserved among all ephrinA proteins at that position. Instead of bulky and hydrophobic Leu124 or tyrosine in ephrinB molecules, ephrinA proteins have smaller serine or threonine at this position (Figure 5). Overall, the D−E and J−K loops of Eph kinases form a small hydrophobic channel, which accommodates bulky nonpolar residues in ephrinB molecules and smaller residues in ephrinA molecules. The higher interaction energies computed for DRY probe (Figure 2) are also consistent with the hydrophobic nature of Eph−ephrin interaction.

```
        36      52       57 61  63 65 68   101    109 155 157          166  190 192 194 196
B2    TATAE EEVS Y M T R Q SVR P F QVDLGGRV KI G C S I
B1    TATAE EEVS Y L T R Q TVR P F QVDFRGRL KV G C I L
B3    WVTSE EEVS Y M P R Q TVR P F RLDAG-RV -- G C S I
B4    LETAD EELS L Q S R E TML P T KRPGAEAT KV G C A L
B5    GETSE EEVS R E Q R Q SVR P F QVDRTGKV RM G C S V
A1    KAQGE SEQQ I - P Y Q TVR P F IRDLASGS KL G C A V
A2    AAGGE DLMQ I - P Y S TVR P T SSDFEARH KL G C A L
A3    TIQGE EEIS V Y P R Q TLR P F QMDLGDRI KL G C A V
A4    SVQGE EEVS M N P R Q TLR P F QVDIGDRI KL G C A V
A5    TVMGD EEIG V Y P H Q TLR P F ELDLGDRV KL G C A V
A6    TVMGE DAIT M N P H Q TLR P F QMDLGDRI KL G C A V
A7    AQQTE EEIS L Y P R Q TLR P F QGDLGERK KL G C A V
A8    TIHGD DSIN V F P H Q TLR P F GADLGFRR KL G C A L
          : .              . ::   *              :: * *    :

        60 62 100  106 109       116 118 120
b2    K D PL R QDVKFT K Q FSPNLWGLE
b1    K D VL R QEIRFT K Q FSPNYMGLE
b3    R D LL R LDLRFT K Q YSPNLWGHE
a1    Y D VR R GPEKLS K Q FTPFTLGKE
a2    Y D KR R GPLKFS K Q FTPFSLGFE
a3    Y D KR R SPIKFS K Q YSAFSLGYE
a4    Y D KR L GHVQFS K Q FTPFSLGFE
a5    Y D KR R GPLKFS K Q FTPFSLGFE
        *          ::: * * ::.
```

Figure 5. Multiple sequence alignment of amino acids directly involved in binding as detected by DIMPLOT. Eph kinases are in uppercase, ephrins − in lowercase. The numeration is that of EphB2 and ephrinB2, respectively.

The meaningful separation between EphA and EphB classes is obtained by carbonyl O probe (Figure 4B). The negative energy contour (in cyan) represents a favorable polar

interaction for EphB kinases targeted by amide group of Asn123 conserved in ephrinB ligands (Figure 5). This asparagine forms a H-bond with Ser109 and nonbonding interaction with Ser194. Respectively this implies that a nonpolar interaction at this position would favor EphA kinases. Surprisingly, ephrinA ligands have a conserved bulky phenylalanine. To accommodate this phenylalanine there is a compensatory change at position 194 to alanine conserved in EphA proteins (Figure 5). It is also interesting that N1 probe suggests a favorable polar interaction for EphA1 and EphB4 kinases at position 158 (not shown), which have arginine there, whereas other kinases have a hydrophobic residue (Figure 5). Thus, CPCA analysis was able to identify several structural features that can differentiate between Eph clusters and some outliers.

DISCUSSION

In this study the 3D-models of ephrins and Eph kinases obtained by homology modeling were subjected to chemometrical analysis using GRID/CPCA approach. In previous works involving this method only receptor proteins were investigated,[13,14] although the method can also be expanded for the studies of ligand proteins. Here, the classification plots were extracted by CPCA from the multivariate data characterizing the binding sites of two interacting proteins, receptors as well as ligands. Both in ligands and receptors we observed two clusters corresponding to A and B subfamilies, respectively. The Eph kinases with unique interaction profiles, like EphA1 and EphB4, were detected as the outliers on the score plot. It was observed elsewhere that this method is very sensitive to conformational changes and is able to distinguish even between bound and unbound protein forms.[16] The ephrin−Eph kinase interactions are best described by induced fit mechanism. The D−E and J−K loops in Eph kinase that form the binding cleft for ephrin G−H loop are unstructured in free receptor but become ordered upon ligand binding. The necessary computational assumption of keeping the protein structure rigid does not reflect the flexibility of the protein interfaces and their structural adaptability. This structural plasticity may be the reason EphB3 kinase interacts with ephrins the same way as other EphB kinases despite the three amino acid deletion in its G−H loop. The homology models, on the other hand, provide a rigid protein structure that is biased by the template protein (EphB2). This explains why we were unable to detect unique place of EphA4 kinase. Although the side-chain conformations can be accounted for in molecular dynamics simulations, the hinge bending movements involved in induced fit docking occur in a much slower scale in order for molecular dynamics to be useful for conformation sampling. The interaction pattern for EphB5 is unknown at the present time, but from the score plots it can be predicted that it would behave essentially like EphB1.

Although the differential pseudofield plots do not provide a drastically different interpretation for individual protein scores, they can explain the structural features different between protein clusters.[14] This approach was able to identify the regions important for ephrin−Eph kinase interactions and regions preferential for one subfamily with respect to the other. The obtained interaction energy contours, which represent selectivity regions for a specific probe, provide the 3D pharmacophore model of the binding site. They can be used as 3D constraints for virtual screening of small molecule databases. We have also identified several amino acids that distinguish EphA1 and EphB4 outliers from other proteins their subfamilies. The role of these amino acids can be probed by site-directed mutagenesis.

As was mentioned above, the difference in charge creates an attractive long-range electrostatic force between ephrins and Eph kinases. This electrostatic force steers two proteins toward each other and aligns ligands in correct orientation with respect to receptors. Despite that, the short-range forces responsible for complex formation are mostly hydrophobic in nature, as can be observed from higher contribution of DRY probe to the interaction. The steric complementarity, described by C3 probe, is also important as bulky amino acids in ligand of one class interact with small amino acids in the receptor of the same class and vice versa.[19]

CONCLUSIONS

In this work we have attempted to determine the mechanisms of ephrin recognition by Eph kinases, although the precise knowledge would require high-resolution crystal structures of the proper complexes. To overcome the gap between available sequence and structure information we have applied GRID/CPCA chemometrical analysis toward the structures obtained via homology modeling. Thus, the characteristic structural features were determined from the analysis of binding site descriptors and not from the sequence alignment. Despite inherent limitations of the homology models we were able to successfully reproduce general patterns in ephrin−Eph kinase interactions and derive the pharmacophore model of the binding site, which can be used in ligand design.

ACKNOWLEDGMENT

This work is supported by grants to B.W. from the Department of the Army (DAMD17-99-1-9019), NIH (CA96533, CA92259), and CaP CURE. E.M. is a recipient of postdoctoral fellowship from Diabetes Association of Greater Cleveland (#F-2002-2).

REFERENCES AND NOTES

(1) Blume-Jensen, P.; Hunter, T. Oncogenic kinase signaling. *Nature* **2001**, *411*, 355−365
(2) Eph Nomenclature Committee. Unified nomenclature for Eph family receptors and their ligands, the Ephrins. *Cell* **1997**, *90*, 403−404.
(3) Kalo, M. S.; Pasquale, E. B. Signal transfer by Eph receptors. *Cell Tissue Res.* **1999**, *298*, 1−9.
(4) Holland, S. J.; Gale, N. W.; Mbamalu, G.; Yancopoulos, G. D.; Henkemeyer, M.; Pawson, T. Bidirectional signaling through the EPH-family receptor Nuk and its transmembrane ligands. *Nature* **1996**, *383*, 722−725.
(5) Lu, Q.; Sun, E. E.; Klein, R. S.; Flanagan, J. G. ephrin-B reverse signaling is mediated by a novel PDZ-RGS protein and selectively inhibits G protein-coupled chemoattraction. *Cell* **2001**, *105*, 69−79.
(6) Kullander, K.; Klein, R. Mechanisms and functions of Eph and ephrin signaling. *Nat. Rev. Mol. Cell. Biol.* **2002**, *3*, 475−486.
(7) Dodelet, V. C.; Pasquale, E. B. Eph receptors and ephrin ligands: embryogenesis to tumorigenesis. *Oncogene* **2000**, *19*, 5614−5619.
(8) Cheng, N.; Brantley, D. M.; Chen, J. The ephrins and Eph receptors in angiogenesis. *Cytokine Growth Facto. Rev.* **2002**, *13*, 75−85.
(9) Wilkinson, D. G. Multiple roles of EPH receptors and ephrins in neural development. *Nat. Rev. Neurosci.* **2001**, *2*, 155−164.
(10) Lu, Q.; Sun, E. E.; Klein, R. S.; Flanagan, J. G. Diverse roles for the Eph family of receptor tyrosine kinases in carcinogenesis. *Microsc. Res. Technol.* **2002**, *59*, 58−67.

(11) Miao, H.; Burnett, E.; Kinch, M. S.; Simon, E.; Wang, B. EphA2 Kinase Activation Suppresses Integrin Function and Causes Focal Adhesion Kinase Dephosphorylation. *Nat. Cell Biol.* **2000**, *2*, 62−69.
(12) Miao, H.; Wei, B. R.; Peehl, D. M.; Li, Q.; Alexandrou, T.; Schelling, J. R.; Rhim, J. S.; Sedor, J. R.; Burnett, E.; Wang, B. Activation of EphA receptor tyrosine kinase inhibits the Ras/MAPK pathway. *Nat. Cell Biol.* **2001**, *3*, 527−530.
(13) Kastenholz, M. A.; Pastor, M.; Cruciani, G.; Haaksma, E. E.; Fox, T. GRID/CPCA: a new computational tool to design selective ligands. *J. Med. Chem.* **2000**, *43*, 3033−3044.
(14) Ridderstrom, M.; Zamora, I.; Fjellstrom, O.; Andersson, T. B. Analysis of selective regions in the active sites of human cytochromes P450, 2C8, 2C9, 2C18, and 2C19 homology models using GRID/CPCA. *J. Med. Chem.* **2001**, *44*, 4072−4081.
(15) Terp, G. E.; Cruciani, G.; Christensen, I. T.; Jorgensen, F. S. Structural differences of matrix metalloproteinases with potential implications for inhibitor selectivity examined by the GRID/CPCA approach. *J. Med. Chem.* **2002**, *45*, 2675−2684.
(16) Naumann, T.; Matter, H. Structural classification of protein kinases using 3D molecular interaction field analysis of their ligand binding sites: target family landscapes. *J. Med. Chem.* **2002**, *45*, 2366−2378.
(17) Toth, J.; Cutforth, T.; Gelinas, A. D.; Bethoney, K. A.; Bard, J.; Harrison, C. J. Crystal structure of an ephrin ectodomain. *Dev. Cell.* **2001**, *1*, 83−92.
(18) Himanen, J. P.; Henkemeyer, M.; Nikolov, D. B. Crystal structure of the ligand-binding domain of the receptor tyrosine kinase EphB2. *Nature* **1998**, *396*, 486-491.
(19) Himanen, J. P.; Rajashankar, K. R.; Lackmann, M.; Cowan, C. A.; Henkemeyer, M.; Nikolov, D. B. Crystal structure of an Eph receptor-ephrin complex. *Nature* **2001**, *414*, 933−938.
(20) Al-Lazikani B, Jung J, Xiang Z, Honig B. Protein structure prediction. *Curr. Opin. Chem. Biol.* **2001**, *5*, 51−56.
(21) Marti-Renom, M. A.; Stuart, A.; Fiser, A.; Anchez, R. S.; Melo, F.; Sali, A. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291−325.
(22) Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779−815.
(23) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.
(24) *GRID*, version 20; Molecular Discovery Ltd., West Way House: Elms Parade, Oxford, UK, 2002.
(25) Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37−52.
(26) Westerhuis, J. A.; Kourti, T.; Macgregor, J. F. Analysis of multiblock and hierarchical PCA and PLS models. *J. Chemom.* **1998**, *12*, 301−321.
(27) Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating Optimal Linear PLS Estimations (GOLPE): An Advanced Chemometric Tool for Handling 3DQSAR Problems. *Quant. Struct.-Act. Relat.* **1993**, *12*, 9−20.
(28) *GOLPE 4.5*, *Multivariate Infometric Analysis*; Srl: Viale die Castagni, 16 Perugia, Italy, 1999.
(29) Marti-Renom, M. A.; Yerkovich,; Sali, A. Comparative protein structure prediction. *Curr. Protocols Prot. Sci.* **2002**, 2.9.1−2.9.22
(30) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283−291.
(31) Morris, A. L.; MacArthur, M. W.; Hutchinson, E. G.; Thornton, J. M. Stereochemical quality of protein structure coordinates. *Proteins* **1992**, *12*, 345−364
(32) Guex, N.; Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **1997**, *18*, 2714−2723.
(33) McDonald I. K., Thornton J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **1994**, *238*, 777−793.
(34) Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: A program to generate schematic diagrams of protein−ligand interactions. *Prot. Eng.* **1995**, *8*, 127−134.
(35) Jeanmougin, F.; Thompson, J. D.; Gouy, M.; Higgins, D. G.; Gibson, T. J. Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* **1998**, *23*, 403−405.
(36) Fiser, A.; Do, R. K.; Sali A. Modeling of loops in protein structures. *Prot. Sci.* **2000**, *9*, 1753−1773.
(37) Cline, M. S.; Karplus, K.; Lathrop, R. H.; Smith, T. F.; Rogers, R. G.; Haussler, D. Information-theoretic dissection of pairwise contact potentials. *Proteins* **2002**, *1*, *49*, 7−14.

CI0256586